

Clinton's horizon beyond the Pacific

Last week's folksy gathering off Seattle should not be mistaken for the country fair it has been represented as. Clinton's immediate goals (and success) notwithstanding, it marks the beginning of an historic process.

It is now more than 20 years since the late Dr Herman Kahn began preaching in the United States the importance of what is called the 'Pacific rim'. Then, he had little evidence to support his case: the productivity of Hong Kong and Taiwan in manufacturing industry, the remarkable recovery of Japan from the Second World War, the emergence of Singapore as a trading centre and the signs, already abundant, that South Korea had set its sights on bettering Japan. From before the Bush presidency, the United States has been alternately angry with and fearful of Japan for its success in the black arts of high-quality mass production which, when technology was simpler, made the Remington rifle, the Singer sewing-machine and the Model-T Ford the symbols of US prowess. Now, uncomfortably, there is China just over the horizon. Kahn had spotted that as well, but the speed of China's economic growth would have surprised him. President Bill Clinton was wise to dig out a lumberjack shirt and hie off to the Pacific northwest a week ago.

Indeed, it was also clever of him, and for several reasons. First, this was not the first meeting of its kind, as may have been inferred from words put about, but the fourth in a series of talking-shops between Pacific governments which the United States has not previously invested with importance. Second, on this occasion, the United States played host, even if offshore. Third, the event was so managed that those who felt left out (including India and Pakistan) were numerous (or at least vocal) enough to give those invited a sense of belonging to an exclusive club. But the real cleverness is that Clinton has left everybody guessing what his motives were. Is he really after a Pacific Trading Area or merely suggesting to the countries of Western Europe what may happen if they do not sign some version of the latest instalment of the General Agreement on Tariffs and Trade (the Uruguay round of GATT) by mid-December?

Even so, the meeting off Seattle will not be the last. Little of substance may have been agreed, but the precedent is bound to set a trend. Those invited will want to meet again, perhaps in more accessible and even comfortable surroundings, while the United States and China each have a need of some multilateral way of talking to each other that does not compromise their principles. What better setting than a talking-shop about technology and trade, in which all parties have a vital interest, that may also serve as a vehicle for an accommodation between the only two credible superpowers of the next decade (beginning in 2001) and may avoid the polarization of the past half-century? Clinton may have

played a flesh-pressing politician's hunch in travelling to Seattle, but there will be a next time (somewhere else) and he will then have serious business on his mind.

What will it be? China's relationship with the nonproliferation regime (see *Nature* **366**, 189; 1993) must certainly be settled by the end of next year. The middling-term agenda is complicated by the rapid souring (early in the 1970s) of the Nixon-Kissinger *rapprochement* with China, whose chief effect (in China's eyes) has been to leave a cohort of bright students stranded in the United States.

The best way forward for the next decade's two superpowers would be to put technology and trade on the footing that strategic weapons enjoyed at the beginning of the Cold War. Then, from early in the 1960s, the Soviet Union and the United States began talking to each other informally about the technology of each other's nuclear weapons. Only when there was a sufficient mutual understanding was it possible, early in the 1970s, for the Soviet Union and the United States to contemplate an agreement on strategic arms control — the premature START-I agreement signed by President Richard M. Nixon at Vladivostok in 1972 (but never ratified). Of course, the same tortuous process will be necessary now in dealings with China, but the agenda should be broader. The past few years have shown that questions such as the expropriation of foreign companies and the exploitation of intellectual property rights can be almost as divisive as the threat of nuclear war. □

Tea-leaves tell nothing

The Council for Science and Technology is not likely to meet the needs of the British science community.

LITTLE can be gleaned about the British government's intentions towards research from the announcement last week that the Director-General of Research Councils (no doubt 'DGRC' in civil service minutes) will be Sir John Cadogan and of the names of the part-time (but paid) chairmen of the three research councils for which there is at present no incumbent. The government is doing what it has (since the early summer) said it would: splitting the Science and Engineering Research Council into two unequal parts (called the Particle Physics and Astronomy Research Council, PPARC, and the Engineering and Physical Sciences Research Council) and providing each with a structure that



resembles that of a private corporation, with a chairman (responsible for strategy) and a chief executive (tactically inclined). It will be easier to guess whether these structures will work (in the sense of being free from personal animosities) when the two missing executives have been identified. But even then it will not be clear whether the new arrangement will be good or bad for research.

Much will depend on Cadogan, whose role (on the corporate analogy) is not clear. Will he stand in relation to the research council chairmen as would the chairman of a holding company to the chairmen of its subsidiaries? Or will he be the equivalent of the principal (and only) shareholder in each of the research council corporations, free to demand an audience with the chairman if he believes funds are being misused or to abuse the chief executive if he considers that too much is being spent frivolously, say on Christmas parties for the staff? On those questions, the government has not been clear; indeed, it appears to have overlooked them altogether. Beyond the duration of his appointment, Cadogan will determine the balance between basic and applied research in Britain. It would be comforting if his undoubted eloquence were matched by more explicit evidence of reflectiveness.

For now-tiny Britain, what will matter most in the years ahead is how literally the rubric of 'wealth creation' (and its extension to 'environment' and 'quality of life') will be understood. Astronomers, of course, will still be free to ask (the PPARC) for, say, funds 'to see whether there are microsecond pulsars', but what will happen to the grant application from a person who believes he is within an ace of understanding the mechanism of cell division? Will prudence require a few thousand extra words about the link between uncontrolled cell division and cancer? And will continuation grants then be turned down if there are no obvious therapeutic applications? And what will happen to palaeontology (except of oil-bearing strata) that cannot masquerade as relevant to biodiversity? We shall not know the answers until the new system is in place next year. The best hope is that DGRC and his fellow chairmen are by then persuaded that they should follow a learning curve of some sort, which is certainly the case.

Meanwhile, the British government has blotted its copy-book on at least one score — the independence (or, rather, the opposite) of its new advisory council, the Council for Science and Technology. No doubt by accident rather than design, two of the new research council chairmen are also members of the advisory committee. Their four fellow-chairmen (and rivals in the inevitable competition for shares of the annual budget) are not. Does that mean that the new council will never discuss the distribution of funds between the research councils? Or that the government will add the four missing chairmen to its membership, in which case it will have recreated the Advisory Board for the Research Councils, abolished only in the summer, but with the difference that the chairman is a minister? Either way, the new council is not constituted to be the sounding board for the opinions of the research community that Britain has for too long lacked. □

White House science

Reinventing science advisory apparatus in the White House should be good for science.

EVER since Jerome Wiesner (now president-emeritus of MIT) was an effective science adviser to President John F. Kennedy, the US scientific community has paid keen attention to the state of science-advising in the White House. Of particular concern is whether the president himself actually listens to his science adviser, as Kennedy plainly did. In the intervening years, there have been good times as well as lean.

Now, President Bill Clinton has taken the first step towards fulfilling his promise to take science seriously. He has announced the formation of two potentially significant advisory bodies (see page 393). The first is a cabinet-level National Science and Technology Council to "coordinate" science, space and technology policy. The president himself will chair the council, whose members will include the scientifically literate vice-president, Al Gore, and the head of the White House science policy office, John Gibbons.

Whether Clinton's attempts to coordinate activities will facilitate collaboration among agencies strapped for money or (let us hope not) prove to be nothing more than restrictive supervision remains to be seen.

In addition, the president has warmed the hearts of US science policy makers by recreating PCAST — the President's Committee of Advisors on Science and Technology, which will consist not of government officials but of independent scientists from academic institutions and industry. The blend is meant to foster what Clinton calls "public/private" partnerships which are considered vital to the national economy but full of conflict-of-interest peril to research universities. The latter are spending increasing amounts of time trying to figure out ways to take industry's money without selling their academic souls.

In the Kennedy era, PCAST was viewed with respect because it represented the voice of academic research and served a role in protecting science from too much of what is today called 'targeted' or 'strategic research'.

It is generally held that devotion to basic science is out of place in today's scientific marketplace, where there is greater emphasis than ever on things like meeting national goals (which means helping to increase national revenues). The new PCAST probably must accept the current jargon but it does not necessarily follow that its members (when appointed) will fail to appreciate the importance of basic science.

It is sheer folly to believe that people engaged in basic work are the equivalent of research airheads with no scientifically sensible goal to their experiments. The leap from basic to targeted research may in some measure lie in language rather than reality. The other thing to say is that budget constraints are real. So the challenge will be to understand that basic science does further national goals. It will be up to science adviser Gibbons, who like Wiesner is in the inner circle, to convince the president of that. □

UK clinical geneticists ask for ban on the patenting of human genes

London. The heads of four leading professional organizations of clinical geneticists have written a joint letter to the European Patent Office in Munich, as well as to senior UK government ministers, asking for a prohibition on the patenting of human genes.

Their demand comes at a time when representatives of the 12 member states of the European Union (EU) are discussing a new directive on biotechnology patents. In its current draft form, the directive would ban patents on gene fragments of unknown function; but it would allow patents on genes whose function is known, and whose potential utility — for example in diagnostic tests — can therefore be described.

This position is backed by many biotechnology companies, keen to exploit the commercial potential of new gene diagnostic and therapeutic techniques. It is also supported by the UK Medical Research Council, even though the council announced last month that it would no longer pursue patent applications for gene fragments (see *Nature* 366, 6; 1993).

But it is now being challenged in a statement agreed by three genetics organizations — the Clinical Genetics Society, the Clinical Molecular Genetics Society and the Association of Clinical Cytogeneticists — as well as the Genetics Nurses and Social Workers Association.

Their joint statement disagrees with the patenting of human genes on two grounds. One is the argument that it is morally unacceptable — and legally questionable — to patent an entity found in a natural state in the human body. (The draft EU directive excludes patents on “parts of the human body *per se*”, but, according to Britain’s Department of Trade and Industry, this exclusion would not cover a functioning human gene outside the body).

The second argument is that the ability to patent human genes is already making genetics researchers increasingly reluctant to share information, because this could undermine future patent applications on their work, and that a reluctance to communicate could lead to a delay in the discovery of new information about genetic diseases.

“My concern is that an excessive concentration by scientists on patenting will diminish the transfer of information, slow up the progress of science and delay the flow of useful information from the laboratories into the clinic,” says Martin Bobrow, chairman of the division of medical and molecular genetics at Guy’s and St Thomas’s Medical School in London.

“Once people start getting too paranoid and looking over their shoulders all the

time, it is likely to slow things down rather than speed things up,” says Bobrow, who is also a member of the newly formed Gene Therapy Advisory Committee, and chairman of the committee that drew up the joint statement.

The clinical geneticists’ statement adds substantial weight to a campaign against the patenting of human genes which is already gathering momentum among the charities that provide more than half of the funds used to support medical research projects in British universities.

Some of these charities, faced with the prospects of gaining extra income through patenting discoveries financed by their own funds — and acknowledging that this practice is already followed by other institutions — admit to being tempted to adopt the same

the issue of gene patents. Many have been prompted to do so by the news earlier this year that the Children’s Hospital in Toronto, whose researchers were the first to identify the main genetic mutation for cystic fibrosis, had been demanding royalty agreements from British researchers developing cystic fibrosis screening kits.

David Owen, who as director of technology licensing at the MRC is responsible for the council’s patenting policy, says that the council does not agree with the anti-patenting line being taken by GIG and others. He says that the question of whether genes should be patented has to be seen in the broader context of the practice and motivation of biotechnology companies.

Owen argues that patent protection is essential if companies are to develop diagnostic tests and new therapies for genetic disorders. “How can you give away [the use of] something that you do not own?” he asks. “We take the view that what is everybody’s is nobody’s.”

But there is a growing feeling, for example among British scientists involved in the international Human Genome Project, that little would be lost — and much could be gained — through an international agreement permitting patents on new diagnostic tests and therapies, but not on the genetic information contained in these tests.

The issue of patents on genes is likely to be raised in a debate scheduled to take place in the House of Lords today (Thursday, 2 December). The Lords’ Select Committee on Science and Technology is already looking at the issue known to be of interest to at least one member, Lord Walton — who is also on the national council of the Muscular Dystrophy Group.

Meanwhile the British debate is being watched carefully by senior officials of the European Commission in Brussels. Similar feelings have surfaced in other European countries; a proposal banning patents on human genes, for example, figures in a recent report to the French prime minister on bioethics. And the whole issue is likely to be debated again in the European Parliament when it discusses the final form of the new EU directive on biotechnology patents.

If the groundswell is strong enough, European countries could agree to include such a ban in the directive. But to be effective (and acceptable to industry) it would also be necessary to persuade other countries, in particular the United States, to adopt the same position. Some are suggesting that this is a task that could eventually be taken on by the Human Genome Organization.

David Dickson

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Pressure for secrecy ‘could slow research on muscular dystrophy’.

stance as the MRC. But most argue that the commercial advantages of patent protection on individual genes are outweighed by other factors.

“The gene is a basic part of the human body,” says Ann Hunt, chair of the Genetic Interest Group (GIG), an umbrella organization of almost a hundred voluntary groups concerned with genetic disorders. GIG has recently issued a position paper in which it argues that patents on genes are against the best interests of those suffering from genetic conditions.

“Quite apart from the ethical and moral objections that we have to the notion of anyone being allowed to own rights over something which is a fundamental part of all of us, we are also very concerned that this trend to patenting will slow down or even stop the development of gene therapy,” says Hunt.

Several individual organizations belonging to GIG, such as the Muscular Dystrophy Group, have already taken a formal stand on

Muscular Dystrophy Group

CERN head pushes claim for 'safe' reactor

Paris. Carlo Rubbia, the director of the European Laboratory for Particle Physics (CERN), has become locked in a dispute with researchers at the Los Alamos National Laboratory in the United States over competing claims for the novelty of techniques which each say is able to produce safe and clean nuclear energy.

Speaking at press conferences in Geneva and Paris last week, Rubbia proposed using proton accelerators to supply neutrons in a new generation of commercial nuclear reactors. These, he said, would eliminate the risk of critical accidents, produce no long-lived highly radioactive waste, and avoid the proliferation of weapons-grade materials such as plutonium.

Rubbia's proposed reactor would use thorium-232, which is plentiful but not readily fissile. Bombarding thorium-232 with neutrons transforms it to fissile uranium-233. The fission of this produces more neutrons which breed further uranium-233 and thorium-232; but it does not produce enough neutrons to sustain the reaction.

His proposal, which is based on sophisticated computer simulations, relies on making the reaction self-sustaining by supplying extra neutrons produced by simultaneously bombarding thorium-232 with protons from an accelerator.

But in a letter last week to the *Tribune de Genève*, Charles Bowman from Los Alamos claimed that his group had previously described Rubbia's concept in detail, in particular in a paper published last year.

In reply, a spokesperson for Rubbia says that he had not published his work because he had decided to file a patent application when he heard that Bowman had already done so. But the proof that the two discoveries are

different lies in the fact that "the two patents are different", says the spokesperson.

Rubbia, who won the Nobel Prize for Physics in 1984 with Simon van der Meer for their discovery of the W and Z particles, maintains that his work is oriented towards energy production, whereas that of Bowman's group is aimed at incinerating nuclear waste. He says that his design therefore uses lower neutron fluxes, and could be developed quickly using existing technology.

Bowman argues, however, that his group also explored low neutron fluxes for energy

production. He says it abandoned this approach because the reactors could not have competed with other designs of nuclear reactors. "Rubbia is re-inventing the bicycle, which was already invented here at Los Alamos," says Bowman. "But it seems he

forgot the tyres," namely the elimination of waste.

Industrialists and nuclear experts have already expressed scepticism about claims by Rubbia and others for accelerator-reactors using thorium in general. Rubbia says, for example, that his reactor is inherently safe because it would always be sub-critical.

But Richard Garwin, IBM Fellow Emeritus at the Thomas J. Watson Research Center in New York, says that criticality is not a major issue, and that the only accident to date — at Chernobyl — was exceptional. The real safety problem facing the industry,

says Garwin, is the prospect of a meltdown following the failure of a reactor's cooling system; the design of Rubbia's proposed reactor does not affect this.

No-one disputes that Rubbia's reactor would produce similar quantities of short-lived (up to 400 years) waste as conventional pressurized water reactors. But, says Garwin, Rubbia's claim that the reactor would produce negligible quantities of long-lived plutonium — 10 grams per tonne of fuel — does not take into account the uranium-238 that Rubbia proposes adding at the beginning of the process.

This spiking is intended to ensure that the uranium-233 could not be used to make weapon's grade material. But Garwin says that the reactor would produce several kilograms of plutonium.

Bowman also claims that Rubbia has failed to state clearly that his reactor would also produce long-lived fission products such as technetium-99 and iodine-129. These are among the most troublesome forms of waste because they dissolve in water and can leak away from geological storage sites.

Nor are industrialists yet banging on Rubbia's door. Jean-Pierre Schwartz, the head of private office of Robert Dautray, chief of the French Atomic Energy Commission (CEA), says that the idea is "worth considering". But he adds that CEA, which is studying accelerators for incinerating waste, has "no immediate plans to explore using accelerators for energy production".

Industry is also cautious. Although particle accelerator-reactors open up new areas for exploration, they will ultimately have to compete with new designs of conventional reactors, and the outcome of this competition will depend on many factors, such as the costs of storing, reprocessing and incinerating wastes.

If uranium prices rise enough, for example, fast-breeder reactors — which are currently out of fashion — may have an advantage over conventional pressurized water reactors. On safety grounds, new designs will also have to compete with the next generation of "very safe" water-cooled reactors which will also be less powerful, so that passive cooling would avoid a meltdown if the cooling system failed.

Undaunted by such arguments, Rubbia says that he intends to pursue the project full-time after he hands over the directorship of CERN to Christopher Llewellyn-Smith on 31 December. "I've been involved in particle physics all my life," he says. "Now I want to do something different."

Bowman says that he welcomes Rubbia's entry into the field. "He is the only person in Europe standing up to say this work is important," he says. And Bowman adds: "I would be happy to collaborate with him."

Declan Butler

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

Photo: Bourque/B. Peller

Carlo Rubbia: seeks 'something different'.

EC pays up for Human Frontiers Program

Strasbourg. The European Commission (EC) has agreed substantially to increase its contribution to the International Human Frontiers Science Program (HFSP), which supports research on brain function and molecular approaches to biology.

The EC's decision follows long negotiations with other participants to persuade it to pay its fair share. Japan, which was the original driving force behind the HFSP, at present pays around 80 per cent of the entire budget. North America and Europe each provide half of the remaining sum.

In principle, the contribution of member states is determined by their gross national product (GNP). Until now, the commission, which represents member states of the European Union (EU) who are not among the seven Economic Summit countries, has paid much less than it should have done. But last week it agreed to increase its contribution to

the 1993-94 budget from US\$200,000 to US\$896,000.

The combined GNP of the eight EU member countries which the commission represents is approximately equal to that of France or Germany. The new agreement brings its contribution into line with that offered by those countries.

Michel Cuénod, secretary-general of the HFSP, has welcomed the increase. But he says that the programme's budget of around US\$37 million this year is still far too low. Although about one-third of the grants applications it receives are given positive reviews, only 12 per cent of grant applications can be funded at present.

Cuénod, who spends most of his time fund-raising, says Japan is prepared to pay much more towards the programme if other countries jointly match its current contribution of US\$30 million.

Alison Abbott

Clinton sets up science and technology council

Washington. President Bill Clinton's new National Science and Technology Council (NSTC) will meet for the first time early in the new year, and will conduct an "across the board review" of US federal research spending during 1994 before delivering advice on the allocation of funds for the following fiscal year.

The council was first proposed in September, largely on the advice of the president's science adviser, John Gibbons (see *Nature* 365, 195; 1993) and was established last week by the president's executive order. It is intended to bring new cohesion to US science policy by bringing together the heads of all government agencies and departments responsible for spending on science, while saving money in the White House by streamlining existing structures.

As an early step towards the latter, the

council will take over the responsibilities of the main such body, the Federal Coordinating Council for Science, Engineering and Technology (FCCSET), as well as those of two other groups, the National Space Council and the National Critical Materials Council.

The full council will be chaired by Clinton and includes all relevant cabinet ministers and presidential advisers, together with the heads of the National Science Foundation, the National Aeronautics and Space Administration and the Environmental Protection Agency.

But meetings will not be regular, only being held as required. According to one White House official this means anything from once a month to twice a year.

The real work of the council will be done by nine coordinating committees, whose areas of responsibility range from health to

international collaboration. These committees will consist of less senior government officials, and will meet much more frequently than the main council. Additional *ad hoc* working groups will be established as required.

Tim Newall, a staff member of the White House's Office of Science and Technology Policy, says that the overall review promised by Clinton will feed into the budget process next year, culminating in the president's budget proposals for the financial year starting in October 1995, which must be ready by January 1995.

Clinton has also established a President's Committee of Advisors on Science and Technology (PCAST) to provide input into the advisory process from the private sector. The committee members and the name of the industrialist who will chair it jointly with Gibbons, will be named early next year.

According to Newall, PCAST will differ from previous panels of its type because it will aim to make direct contact with US industry through active subcommittees, "instead of just meeting once a month in Washington".

Officials say the cabinet-level NSTC is proof of the Clinton administration's commitment to science and technology, and that its responsibilities will mirror those of the powerful National Security Council.

But the existence of the new body by no means assures that it will carry clout. Science and technology priorities in the United States tend to be set independently by government departments and other agencies; their defenders in Congress are likely to oppose any attempt to centralize decision-making.

Colin Macilwain

Saxony rejects appeals court verdict

Munich. The German regional government of Saxony has refused to reinstate a Leipzig professor who has appealed successfully against his dismissal last August for alleged cooperation with Stasi, the former East Germany's secret police.

Hans Joachim Meyer, the *Land's* research minister, has refused to accept a court decision that Wolf-Dietrich Arnold, previously director of the University of Leipzig's Orthopaedic Clinic, should be given his job back. Indeed, Meyer has approved the provisional offer of his job to a new applicant for the position, pending the decision of a higher court.

Arnold, who has the support of most members of his department, has always denied the charge that he used his membership of the Communist party to the disadvantage of the careers of his students and staff (see *Nature* 359, 762; 1992). He claims that he joined the party only five years before reunification to make his professional life easier by increasing his political contacts.

Arnold is one of thousands of academics in the former East Germany who have been dismissed during the reorganization of the universities that followed reunification. Hundreds have appealed against the decisions, and at least 14 appeals in Saxony have been successful. But Meyer has challenged them all, saying that he will continue to fight any victory as a matter of principle.

Arnold's case was the last to be heard. On 15 October, the research minister offered to settle, a move that would have given victory to Arnold after an obligatory four-week 'consideration period' during which either party has the right to change its mind. Shortly before the end of this period, however, Meyer withdrew the offer and announced that he would be appealing to a higher court.

This case comes up next month. Meanwhile — and in defiance of court decisions — Meyer has allowed the University of Leipzig to seek a replacement for Arnold. The selection process has already taken place, and the directorship has been offered, in principle, to a new candidate. Legally the post cannot be definitively offered unless the higher court rules in favour of Meyer.

A spokesman for the ministry said that it could make no comment while the court case was pending. But Arnold believes it is "monstrously unfair" of Meyer to pre-empt a court decision and decide on a successor in this way.

Allison Abbott

Mathematics inspires office of the future



Mathematicians visiting the Isaac Newton Institute for Mathematical Studies in Cambridge, England (left) may be excused for finding themselves stimulated as much by the bricks and mortar as the scientific debate going on within the building.

The institute, opened in June last year, has won the 1993 Office of the Year Award for Innovation sponsored by DuPont (UK) Ltd. The judges described it as "a model for commercial buildings of the future where professional and knowledge [*sic*] workers will be employed in greater numbers".

The focus of the building, designed by Duncan Annand, a local architect, is an open common room where researchers can 'brainstorm' on the blackboards that decorate the walls as they meet over coffee. The institute brings together researchers from different countries and disciplines, usually for a relatively short period of time.

Fiona Gammie

Catalonia wants greater control of institutes

Barcelona. The government of northern Spain's autonomous region of Catalonia has offered to take over the running costs of the 13 institutes in its area of the country's Council for Scientific Research (CSIC), which is currently under financial pressure.

The move is part of a bid by Catalonia to integrate the research programmes of the CSIC institutes, at present funded and controlled by central authorities in Madrid, into its own regional science and technology plans. But these authorities are resisting any significant reduction in their influence.

Run by a centre-right coalition government, the region is seeking to develop a less centralized system for supporting research and development than in the past. Its goal is to move closer to the German model, where individual *Länder* share with central government responsibility for both funding and directing the work of research institutes.

Discussions have been taking place for more than a year between officials from Catalonia's Commission for Universities and Research and those from the Ministry of Education and Science (MEC) — on whom CSIC depends — in Madrid. Both sides confirm that an agreement is likely to be

reached by the end of the year.

CSIC has more than 200 permanent researchers and another hundred visitors and undergraduates in Catalonia. Some of the CSIC institutes in the region are among the best in Spain. In the area of biology, biotechnology and biomedicine, for example, the Centro de Investigación y Desarrollo in Barcelona is CSIC's most productive institute in terms of scientific output, and its third in terms of size.

Three factors lie behind Catalonia's bid to take on increased responsibility for local research activities. The first is a feeling that although the central government has done much in recent years to boost the country's scientific capabilities, it has been less successful in transferring results to industry — and that this can probably be best done on a regional basis. "We are reaching a glut on the supply side. We need a change of focus in order to find ways of ferrying people — and ideas — to the productive sector, because the question now is to satisfy their demands", says Josep Laporte, head of the commission.

As part of its own efforts, for example, the regional government plans to inaugurate

two so-called 'reference centres' later this month — one in biotechnology and the other in food sciences — to act as focal points for research groups in fields being given priority. Three more centres, in new materials, environment and advanced manufacturing, will be created next year. Each will aim to stimulate the transfer of innovative technologies to industry.

The second factor is Catalonia's feelings that its regional needs are not being properly met by Madrid. Joan Albaiges, who resigned as the CSIC's delegate in Catalonia to become the commission's director-general of research, points out that Catalonia now provides as much money as the central government to the support of research and development in the region. Its research plan has a budget of Ptas 6,800 million and will grow 3 per cent next year despite a general cut in local administration of 20 per cent.

But he says that the distribution of funds unfairly favours institutes close to the centre. "The state gives us 12 per cent of its total for R&D and 42 per cent to Madrid. But our scientific production represents 21 per cent of the total and Madrid's only 30 per cent. This is ridiculous," he says.

The third factor is political. Last June, the Socialist party failed to win a majority vote in the general elections. As a result, it has had to rely on the votes of Catalonia's nationalist coalition in order to approve the national budget for 1994, and is likely to require its support in passing other legislation. Agreement on the CSIC centres appears to be a concession that the Socialists have made to ensure this support.

Last year, Catalonia asked Spain's constitutional tribunal to require the central government to devolve responsibility for research (including the work of CSIC institutes) to its autonomous region, but the bid was rejected. "Now, we want to achieve the same through negotiation," says Albaiges.

Elias Fereres, state secretary for universities and research, accepts that the new agreement will mean increased involvement by the local government in the activities of the CSIC institutes. (A similar agreement has already been reached with the region of Andalusia, which includes Seville.) But he denies that the agreement would result in any way in a devolution of powers to Catalonia on research.

"It is just a question of coordination with their plans," says Fereres. He argues that, under the terms of the proposed agreement, CSIC's programmes in Catalonia would be "oriented" by the autonomous government, but would not be directed by it.

"The agreement would not mean reducing the central government's budgetary responsibilities; it is not our intention to reduce our financial support of CSIC", says Fereres. **Luis Angel Fernández Hermana**

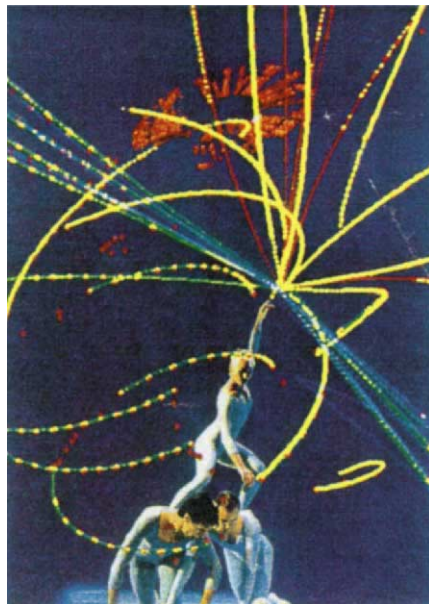
Europe takes its science out to the people

Paris. The European Commission organized a "week for scientific culture" last week to encourage the public to take a greater interest in science. Top prize in meeting this objective must go to the Italians, French, British and Czechs, who targeted the Nintendo generation with a multimedia event in Trieste (Italy) linking science fiction movies and comics, documentaries and celebrities.

CERN packed crowds into its underground particle-accelerator for an open day, and added some culture with a ballet (right) on the Big Bang in Geneva's Grand Casino. The United Kingdom went for smaller numbers, but compensated by targeting the young and impressionable, inviting 350 children for a sleepover at London's Science Museum.

Others interpreted the theme liberally. Spain and Italy organized a conference for businesspeople, scientists, administrators and journalists (but apparently not the public) on saving the Mediterranean. And only France could have held a debate between philosophers, historians and scientists under the title: "European Science in Self Question".

Events with less enigmatic — if more prosaic — titles include a conference in the Netherlands on "The responsibilities of industry in science education", and a debate in Strasbourg organized by the European Science Foundation and enti-



tled "Chemistry without prejudice — a day to reconcile the public with a science which is omnipresent but misunderstood in daily life."

The prize for self-promotion must go to the European Union's four Joint Research Centres (JRC) in Italy, the Netherlands, Germany and Belgium. These invited three journalists from each of the four countries to explore JRCs in the other three.

Declan Butler

Top US panel will map out fresh attack on AIDS

Washington. Donna Shalala, the US Secretary of Health, is to set up a National Task Force on AIDS Drug Development in an attempt to revive waning hopes that a cure can be found for the disease.

According to a statement due to be issued in Washington this week by the Department of Health and Human Services, the 15-strong panel will be chaired by Shalala's assistant secretary, Philip Lee, and charged with "seeking new and innovative approaches to the development of AIDS drugs". It will concentrate on identifying and removing obstacles in the path of drug researchers, the department says.

In announcing the initiative at the National Institutes of Health (NIH) on Tuesday, Shalala was expected to stress the need for new and innovative approaches to the development of drug treatments for HIV and AIDS. At present, officials say, there are no new drug applications for antiretroviral drugs being considered by the Food and Drug Administration (FDA).

Shalala is inviting nominations for panel members. But membership is likely to include most of the prominent people due to accompany her at the announcement, namely White House AIDS coordinator Kristine Gebbie; Anthony Fauci of the National Institute of Allergy and Infectious Diseases (NIAID); Harold Varmus of NIH; David Kessler of the FDA; Roy Vagelos, chairman of the drug company Merck; and Moises Agosto of the National Minority AIDS Council.

The idea of the task force came largely from Kessler. "This is trying to look at the entire continuum, from basic science through to clinical treatment," says Jim O'Hara, an FDA spokesman. "The panel will try to look at this in a seamless fashion, and see where the potholes are on a track."

But the initiative will also be seen as a political response by the Clinton administration to growing disenchantment and despair over the prospects for treating AIDS, particularly in the wake of preliminary results of the Anglo-French Concorde trial earlier this year, which showed that AZT does not prolong patient life.

AIDS activists and researchers had not been informed of the details of the plan as *Nature* went to press. But the initiative is likely to be welcomed as an attempt to address directly the central issue of finding an AIDS cure. "It sounds like they are all trying to put their heads together", says Gary Bonasorte of the Community Research Initiative on AIDS in New York, which has been trying out innovative approaches to HIV treatment. "It's about time."

Colin Macilwain

NSF chief urges support in selling science to the public

Washington. The ending of the Cold War makes it more important than ever for scientists to explain why public support for their work is necessary, according to Neal Lane, the new director of the US National Science Foundation (NSF). But he also warns that congressional pressure to shift resources from basic to strategic science means there is unlikely to be any growth in funding for areas such as astrophysics and astronomy in the near future.

Scientists should do more to sell themselves to the public, to ensure that science retains both public support and the funding that goes with it, Lane said in an interview last week. "We have to do it, as an agency justifying our budget," he says.

"But the wider community of scientists and engineers can help. They have an opportunity to take this question seriously, and to provide the best possible answers." Lane was confirmed in October as director of the NSF, the \$3-billion agency that funds most US university research in the physical sciences, as well as projects in biology, social sciences and education. One of his first tasks has been to oversee the NSF's response, due in January, to sharp criticisms made recently by the Senate appropriations subcommittee responsible for approving the foundation's budget.

The subcommittee is threatening to divert money to other agencies if the foundation does not pay more attention to applied research. It told the NSF to spend at least 60 per cent of its money on so-called 'strategic' research, and explicitly warned it "not to shroud curiosity-driven activities under the rubric of strategic activities".

"My reading of the report is that the committee feels the NSF has not been as responsive as it would like in explaining what we were doing, and in setting priorities," says Lane. "The committee wants the NSF to take more seriously our responsibility in helping the nation with some of the very serious problems that our people face."

Lane says that the Senate is also asking the foundation to be more precise in how it defines strategic research. He cites his own field of atomic molecular optical science as one example of work of a fundamental nature which can still qualify as 'strategic' because it matches the mission of a government agency.



Neal Lane: confident of making an impact.

"We have no intention of stopping funding areas of research like astronomy, astrophysics, cosmology or pure mathematics," Lane says. "Certainly we expect to support healthy programmes in all of those areas." But pressed on the prospects of growth in such fields, Lane strikes a cautious note. "Our country has financial problems which do not imply huge budget growth. So priorities need to be set on how money should be spent," he says.

Lane accepts that the NSF needs to redefine its mission in the wake of the Cold War. Military requirements were only one element of the thinking behind the creation of the agency in 1948, he says. "But the world has changed in such a substantial way that it is important for the question [about its mission] to be asked, and for us to explain why our work is so critically important to the country."

To help to do this, individual scientists and engineers must "reach out into their community", says Lane. "Many are already doing that, but some don't, because it's not quite clear how to get started," he says. Lane promises that the NSF will coordinate attempts to get scientists out into factories, offices and schools.

"Every high technology business in the country is doing what it is doing based on knowledge that came out of scientific research, funded largely by the government," says Lane, who appears comfortable justifying the NSF budget on the basis of industrial competitiveness. Economies such as Japan which historically relied less on pure scientific research "now recognize the importance of basic research, and of having it in their own country," he says.

Lane, who was formerly provost of Rice University in Texas, says he is confident of making an impact as director "because this agency has such a track record of success that it is possible for one person to come in, provide a focus, and enable it to pull together its collective energies".

He claims that the Clinton administration's commitment to science and technology is exemplified by its treatment in the recently completed budget round, which he describes as "really quite good". Lane adds: "I think you'll continue to see that unfold in the coming months and years."

And Lane says he is "very positive" about NSF's move away from its old headquarters, a stone's throw from the White House, to Arlington, in Virginia. "This building has been home to the NSF for a long time, so people understandably have strong feelings about it," he says. "But I don't feel that physics or the NSF are being downgraded by the move."

Colin Macilwain

Science publishing in Russia 'on verge of catastrophe'

Moscow. The bleak prospects facing Russian scientific publishers have become even grimmer in recent months following a series of moves by the country's chief academic publishing company, Nauka Concern. In particular, Nauka has been threatening to close its subsidiary Fizmatlit, which specializes in publishing journals and books in physics and mathematics.

The economic crisis of recent years has faced Fizmatlit with a severe decline in production which has brought it to the verge of bankruptcy. However, Fizmatlit's director, Vitaly Kulyamin, believes that ways can be found out of its current predicament.

Its prospects have been brightened by contacts with the German publishing company Verlag Harry Deutch. Fizmatlit has also signed a potentially lucrative agreement with the Foundation for Charity and

factories to be sold at knockdown prices. The printing-works can no longer produce books, and is now able only to print letter-head and forms.

Nauka Concern is more commonly known to the Russian scientists under its old name of the Nauka Publishing House. It was created in 1964 as an amalgamation of several smaller publishing organizations and printing houses, all specializing in academic literature.

This arrangement fitted in well with the centralized Soviet system. But it became unviable with the arrival of the free market. Given the current economic crisis, scientific publishing in Russia has little chance of becoming financially profitable, and is unlikely to survive without significant outside help.

Kulyamin says that small publishing companies have a better chance of survival because of factors such as higher flexibility and lower costs. Large companies such as Nauka would have considerably more difficulty trying to survive as a scientific publisher. Some observers feel that it is preparing to transform itself into something more profitable — and less scientific.

Nauka's attempt to close down its own publishing subsidiaries has generated protests from the scientific community. "This is yet another way to devastate science in our country", says Samarsky.

The output of large publishing companies in Russia has recently diminished by, on the average, more than two thirds. The losses of scientific publishers are difficult to assess, but are probably even higher. Scientific journals are still subsidized by the state, but scientific book publishers have to struggle on their own, without even any relief from tax.

Furthermore, the Russian Foundation for Basic Research (RFBR) recently decided to cut back its support of scientific publishing, on the basis of what it claims is the gradual disappearance of those in need of such support.

Russians no longer write monographs, textbooks, or popular science literature because they are unable to find funds to support the publication of their work (publishing companies require their authors to at least cover any losses). But even when financing is made available — for example through an RFBR grant — very few people take up such an offer.

Officials at the foundation hope that a new competition for grants will stimulate greater scientific activity. But they acknowledge that the situation of book publishing in Russia is potentially catastrophic.

Vladimir Pokrovsky

Brussels gives bigger role to European Science Foundation

Strasbourg. Paolo Fasella, director-general of the European Union (EU)'s research commission, has welcomed an offer from the European Science Foundation (ESF) to play a greater role in research programmes financed through the commission in Brussels. His remarks followed a formal letter from Antonio Ruberti, the commissioner for research, to the ESF requesting more formal contact between the two bodies.

Fasella told the member organizations of the ESF attending their annual meeting in Strasbourg last week that the commission was "very positive" about using the scientific expertise of the ESF to help to evaluate research programmes, to identify referees for the programmes, and even, in certain cases, to manage some research projects.

The ESF, which runs a variety of scientific activities for its 54 member research organizations in 20 European countries, has been actively seeking stronger links with the commission (see *Nature* 366, 193; 1993). In particular, it wants to establish itself as a formal advisory body to the commission, representing the 'voice of science' in European-level affairs.

Some commission officials are said to have reservations, fearing that they would lose control over decision-making. But Ruberti has long been in favour of closer links with national research organizations. Last May, for example, he invited delegates from Germany, France and the United Kingdom to Brussels for formal talks.

Fasella told the ESF that the research commission needs an "independent, respected body" to advise on its proposals for scientific programmes, and that such a body could "bring regularly to the attention of political decision makers important scientific areas that may be of interest". The ESF fulfils the necessary criteria, he said.

He also acknowledged the difficulty the commission has always had in identifying scientific referees for its programmes, and the much criticized lack of transparency in the process. "There must now be both *sein* and *schein*", he said, implying a commitment to a system that must be both fair and seen to be fair. He foresees the ESF, with its access to national lists of potential referees, playing a fundamental role.

Fasella also said that the commission would in certain circumstances be prepared to hand over the management of commission research projects to the ESF.

But Fasella also warned that the commission will have to be convinced of the cost-effectiveness of any delegation, because it is legally obliged to spend no more than five per cent of a project budget on administration.

Alison Abbott



Culture, under which it plans to publish educational literature aimed at gifted children.

But during a visit to Germany in October, Vladimir Vasiliev, director of Nauka, told Fizmatlit's German partners that Fizmatlit was being closed, and that they should negotiate with him directly over any future agreements.

Only after the German company, backed by several well-known Russian academicians such as Vitaly Ginzburg, Nikita Moiseyev and the mathematician Alexander Samarsky, had expressed strong objections were the liquidation plans suspended.

Kulyamin describes Nauka's threats as a "personal conflict". But Nauka has acted in a similar way towards other parts of the organization. For example, it ordered much of the printing equipment in one of its

Taiwan beckons researchers back from US

Hsinchu, Taiwan. Chinese scientists attracted back from the United States are expected to drive the expansion plans of Taiwan's highly successful science-based industrial park in Hsinchu as it enters its third phase of development.

Over the past ten years, nearly 150 companies from Taiwan, the United States, Europe and elsewhere in Asia have set up shop in the science park, which is administered by the government's National Science Council (NSC). Hundreds of Taiwanese scientists and engineers have returned from the United States to set up businesses.

Most activity so far has been in computer and telecommunications related enterprises. Annual sales are about to exceed NT\$100 billion (US\$3.7 billion). The park, about 60 miles southwest of Taipei, employs 28,000 people, more than 2,000 of whom hold doctorates or masters' degrees in science and engineering. Many of the junior employees have been trained at two universities close to the science park, National Chiao Tung University and National Tsing Hua University.

The nearby Industrial Technology Research Institute (ITRI), funded by the Ministry of Economic Affairs, has provided several spin-off companies and talented researchers. But most of the leaders of companies are what Steve Hsieh, director-general of the park administration, jokingly calls "CEOs" — or "Chinese Entrepreneurs from Overseas" — coming mainly from



Integrated circuit manufacturing in Hsinchu owes much to Silicon Valley.

California's Silicon Valley.

The NSC administration rents out land and buildings in the park and helps companies to cope with the red tape involved in setting up and operating a company in Taiwan. Tax incentives are offered in the first few years of operation, as well as financial support for research and development.

The park administrators also showed foresight in establishing from the outset in 1983 a bilingual school for the children of scientists returning from the United States and elsewhere in the West. In this regard, Hsinchu is far ahead of Tsukuba science city in Japan, which after 20 years still has only a tiny privately run international school.

The NSC provides competitive grants of

up to NT\$5 million to fund up to 50 per cent of research and development projects carried out by companies. About a quarter of the companies in the park receive such grants each year. This year, the government introduced new incentives to encourage research on 'strategic' products that will increase Taiwan's competitiveness in world markets. Companies have so far applied for NT\$1.4 billion for 21 strategic-product development plans.

About 200 hectares of land have been acquired next to the existing 380 hectares of the park for the next phase of development. The first companies are expected to set up shop next year.

It has taken more than two years to win approval for the next phase from the local people. "They now realize the value of the park in attracting other small industries and they want to be located close to the development", says Angela Yao, promotion representative for the park administration. The park intends the next phase to concentrate on biotechnology, in particular the production of vaccines for diseases such as hepatitis that are common in Taiwan, as well as medical equipment and food products.

It is not yet clear where the necessary researchers will come from. Taiwan's domestic pharmaceutical companies are weak and have no significant research and development base. In this regard the situation is different from that for integrated circuits, where ITRI provided a ready source of talent and several spin-off companies.

But there is a huge reservoir of bioscientists of Chinese origin in the United States, a pool of talent which the park will no doubt try to tap. Cheng-Wen Wu, director of the Institute of Biomedical Sciences of Academia Sinica in Taipei — who himself returned from the United States — estimates there may be as many as 5,000 such Chinese bioscientists in the United States.

David Swinbanks

Chinese institute appoints new director

Peking. After the first ever international competition for a top scientific appointment (see *Nature* 364, 474; 1993), the Chinese Academy of Sciences has announced that the new director of the Institute of Geophysics is to be Xu Wen-Yue, a research professor at the institute.

Thirty candidates applied for the post when it was advertised earlier this year, a response to growing pressure on the country's research institutions to reform their rigid personnel policies. Xu Wen-Yue was selected from a short-list of four, after consultation with the staff at the institute. **You Qin Li**

Synchrotron goes in search of users

Hsinchu, Taiwan. As the administrators of Taiwan's science park turn their attention to biotechnology (see above), the country's physicists are expressing pride in their first synchrotron, now entering the final stages of commissioning at the new Synchrotron Radiation Research Center in the park.

The achievement is impressive. A country with no experience in building synchrotrons has constructed a sophisticated third-generation machine and reached its operational targets a few months after tests began last April. In August, the synchrotron exceeded its target of a beam current of 200 milliamperes, reaching 226 mA, and achieved a beam life of more than five hours.

It is Asia's first such machine. There are more powerful third-generation synchrotrons in the United States and Europe, but Taiwan's has begun operation well before South Korea and Japan bring their first third-generation machines on line.

The big question now is who is going to use the 1.3 GeV machine. The vast circular hall surrounding the US\$110-million synchrotron is largely empty. There is room for

44 beam lines. But at present only three are being operated by a handful of researchers.

None of the integrated circuit manufacturers in the park has expressed interest in using the facility. And, even though training programmes are under way, there are only a limited number of scientists in Taiwan capable of using such machines.

Yuen-Chung Liu, director of the centre, is confident that scientists from overseas will come to use the machine once commissioning has been completed over the next few months, and the capabilities of the machine have become clear.

For example, Liu — who has spent much of his life in Japan — expects scientists building the Spring-8 synchrotron in western Japan to use Hsinchu until their more powerful machine comes on line later this decade.

But many Taiwanese question the value of this investment at a time when funding for research is being squeezed by government spending on new public construction works. To prove its worth, the centre hopes it will be able to draw in users quickly from overseas, as well as from Taiwan itself. **D. S.**

Effects on health of mustard gas

SIR — Graham S. Pearson, in his review of *Veterans at Risk: The Health Effects of Mustard Gas and Lewisite*¹ dismisses suggestions that US veterans exposed to small amounts (2 drops) of mustard gas on the skin should be identified. He points out that, in the tests, one drop of mustard gas was removed immediately with a decontaminant and the second two minutes later.

Because a long-term health risk has not been identified, Pearson asks, why should those who took part in the tests be worried by being contacted at this stage? The 4,000 US servicemen exposed to higher concentrations of chemical warfare agents are, as Pearson says, the ones most at risk, and, given the dearth of information on the effect of poison gas on human beings, this cohort must be the one to follow in the first instance.

Mortality records of British and US servicemen exposed to mustard gas suggest an increased incidence of lung cancer², although smoking was a confounding factor in the British study. Chronic bronchitis was also increased among UK veterans exposed to mustard gas, but this, too, is difficult to evaluate because of uncertainty about smoking habits³. Civilians involved in the manufacture of mustard gas have a high incidence of lung cancer and respiratory cancer in general⁴.

It may now be too late to confirm whether lung or respiratory cancer is increased in servicemen exposed to mustard gas on the battlefield; the First World War was the last conflict in which troops were caught either without, or with inappropriate, gas masks. The exceptions, of course, are the Iranians injured by mustard gas between 1983 and 1988 in the Iran–Iraq war; many of them had ill-fitting masks. Kurdish civilians in Iraq had no masks at all when they were gassed in 1988.

There is also the outstanding question about the long-term effects of exposure to nerve agents. It is known that Britain's Chemical Defence Establishment at Porton Down carried out studies to develop a prophylactic treatment against nerve gas (and against the agent soman in particular). Porton's work on the prophylactic could be a real life-saver. Conventional treatments for dealing with other organophosphate nerve agents are much less effective against soman.

Porton's studies were said to involve 300 men consuming tablets of the carbamate pyridostigmine bromide followed by exposure to the organophosphate nerve agent sarin. Total exposure would (in theory) inhibit 60 per cent of acetylcholinesterase activity in serum, whereas in practice the combined treatments resulted in about a 50 per inhibition⁵. It is the

inhibition of this enzyme at the neuromuscular junction that causes many of the acute symptoms in those poisoned by organophosphates.

There are outstanding questions about the long-term health effects of acute (and chronic) exposure to both organophosphates and pyridostigmine. Porton's 'volunteers' might prove to be a useful group to follow up for some answers. There is documented evidence of exposure, blood chemistry and, presumably, a register that could be used to trace individuals.

Alastair Hay

*Department of Chemical Pathology,
Old Medical School,
University of Leeds,
Leeds LS2 9JT, UK*

1. *Nature* **365**, 218 (1993).
2. *IARC Monographs* **9**, 181–192 (1975).
3. Case, R. A. M. & Lee, M. J. *Brit. J. Prev. Soc. Med.* **9**, 62–72 (1955).
4. Wada, S. *et al. Lancet*, **1**, 1161–1163 (1968).
5. Gall, D. *Fundam. Appl. Toxicol.* **1**, 214–216 (1981).

SIR — Graham S. Pearson, who is associated with the Chemical and Biological Defence Establishment at Porton Down, complains that the Institute of Medicine (IOM)'s study committee included no members from the US Department of Defense (DOD) and only a few members of the IOM or National Academy of Sciences (NAS). But the policy of the IOM and NAS is to appoint committee members on the basis of expertise; the inclusion of four IOM members and one NAS member is not unusual; some committees have more, some fewer. It is also policy to exclude individuals who, by their relationship with specific government agencies, may have conflicts of interest on the subject of a study. That disqualified members of the armed services or employees of the DOD from serving on the committee, but the views of such individuals were sought (and are detailed in Appendix A of the report). The committee was also given a presentation by a person who had been actively involved in chemical testing programmes during the Second World War.

Pearson accuses the committee of ignoring the standards of the time and forgetting that chemical warfare was a reasonable expectation in the Second World War. But the committee did consider this question, but could not ignore clear evidence of abuse. For example, Chapter 4 says that the committee "believes that the... investigators involved... were convinced of the likelihood of great numbers of gas casualties". Yet the responsibilities of physicians then and now are to "do no harm". Even War Department permissions to use servicemen in these experiments included that

condition, yet many were injured from exposures equivalent to First World War gas attacks.

The treatment of the "volunteers" was documented in the IOM report by examination of the military's experimental records. This treatment included outright threats if subjects refused repeated gas chamber trials, sometimes despite the presence of symptoms of gas poisoning described by military physicians before the Second World War. One could argue that wartime exigencies should modulate judgement of these tactics, but (as the report documents) such practices were continued by the military in the United States well into the 1970s. Also, although the Nuremberg Code did not emerge until after the war, a central thrust of the trials that produced the code was that its basic principles governed conduct before and during the war.

From incomplete records, the IOM committee knew of *at least* (rather than *only*) 4,000 servicemen involved in the high-exposure chamber and field tests. The military historian Rexmond Cochrane (cited in the report) reported 60,000 total participants in chamber, field and patch tests. The committee's recommendations regarding illnesses such as lung cancer and emphysema clearly stated that these would occur only in chamber and field-test participants, not among those only exposed in patch tests. The US Department of Veterans Affairs, however, accepts skin cancer in the scars of patch-test participants as causally related to mustard gas.

Pearson states that this committee's standards would argue for "compensation" of nonsmoking military personnel or students in chemistry laboratories for the effects of passive smoking and benzene despite the fact that the effects of these substances were not known at the time. This analogy is inappropriate, because military research, some published in open scientific literature before the Second World War, documented long-term health effects in gas victims from the First World War — chiefly bronchitis, emphysema, chronic asthma, corneal opacities and keratitis. The records show that no attempts were made by the military to warn the subjects of the possibility of these debilitating effects. In fact, the subjects were forbidden on threat of prosecution to discuss their experiences, even with their personal physicians. That oath of secrecy was finally lifted on 10 March 1993. In the report's recommendations, the term "compensation" is never used. Rather, the Departments of Defense and of Veterans Affairs were asked to try to find affected veterans still living and to evaluate and treat medical problems caused by the exposures they received.

Finally, Pearson accuses the committee of unreasonably raising "alarm and anxie-

ty" among US servicemen and their families. A full reading of the report would show that the committee struggled with this idea throughout. Differences in exposure levels between the test types were explained in detail and an expert in risk communication was consulted in an attempt to reduce potential anxieties and fears. These efforts were balanced against the desire voiced by veterans to know what had happened to them and to end decades-long doubts about the causes of their health problems. The committee clearly states that not all the veterans' questions could be answered on the basis of rigorous scientific analysis. Yet, the veterans who have contacted the committee (now well over the 250 cited in the report) seem most grateful that their experience, so long ignored and denied, was finally affirmed.

David P. Rall

(Committee Chair)

Constance M. Pechura

(Study Director)

*Institute of Medicine (F03036),
2101 Constitution Avenue, NW,
Washington, DC 20418, USA*

Fidia and neuroscience

SIR — In the press coverage concerning the falterings of the Fidia Pharmaceutical Corporation in Italy (*Nature* **364**, 562; 1993) and the financial fall-out threatening the viability of the Fidia Georgetown Institute for the Neurosciences (FGIN), the most important dimension seems to have been overlooked.

Lost somewhere between enumerations of broken promises and legal battles are the significant human resources that define the FGIN. This institute, under the directorship of Professor Erminio Costa, is comprised of our fellow faculty and scientific colleagues, who, since the institute's inception in 1985, have dramatically enriched the intellectual and academic environment at Georgetown University. The invaluable contributions of these eminent scientists should not be reduced to a price tag or a percentage of a budget.

As faculty members of various departments at Georgetown, we have benefited in many ways from our collaboration with members of FGIN. This special relationship has attracted students and faculty. The research conducted by this group has enhanced our research programmes and resulted in the initiation of several new projects funded by the National Institutes of Health. FGIN has trained students and scientists from less wealthy countries, inspiring us to follow suit. Dollars cannot measure what we have gained by the growth and development of the FGIN in our midst. Nor can money mea-

sure what we stand to lose if this group of neuroscientists is not given the support and encouragement necessary for them to continue at our university.

Karen Gale

on behalf of 33 members of the faculty
*Georgetown University Medical Center,
3900 Reservoir Road NW,
Washington, DC 20007-2195, USA*

SIR — Your brief account of the troubles of the Italian pharmaceutical company Fidia was unfair with respect to the support it has given to basic neuroscience. I do not wish to defend the actions of the Italian government, and clearly Fidia Pharmaceuticals has made managerial mistakes for which they are paying, but to refer to the Fidia-sponsored programme of scientific conferences, travel grants, prizes to established and young scientists, training of postdoctoral fellows (as in my case) and the establishment of laboratories in Washington DC as "eyebrow-raising largesse" and worse still "a rat to be smelled out" is not nice. Indeed Fidia has spent more money (as percentage of turnover) on research and development than any other European pharmaceutical company, largely supporting basic research (see *Nature* **361**, 765–768; 1993). Fidia's fall-out should be seen with sadness rather than derision.

Stefano Casalotti

*Neuro-behavioural Biology Center,
Mahidol University, Salaya,
Nakorn Pathom 73170,
Thailand*

SIR — We are concerned that your article may give the impression that, following a lawsuit by Georgetown University against Fidia over a reimbursement due in 1995 for the construction of a research building, the university has decided to retaliate by closing the Institute of Neuroscience (FGIN). We believe that such a move would greatly damage relations between universities and industry.

Although your report states that FGIN represents only 10 per cent of Georgetown's spending on neuroscience, the importance and impact of the contributions made by FGIN dwarf any other results that the university may have obtained with the remaining 90 per cent of neuroscience research spending. One might conclude that the size of investment does not always guarantee the quality of science, but it is the latter that counts.

The FGIN achievements have been assessed by an *ad hoc* committee selected last year by the university to carry out a periodical in-depth review of the institute's scientific activities. The committee, consisting of A. J. Aguayo (president), S. F. Heinemann, K. Fuke, H. Mohler and E. M. Johnson, prepared a positive and laudatory report. Moreover, in a survey in *Nature* of industry-sponsored activities in

US universities (**361**, 765–768; 1993) the institute's activities were evaluated positively. Surprisingly, your article states that FGIN will not survive, which contrasts with a report on the same topic that appeared almost simultaneously in *The Lancet* (**14**, 625; 1993).

The report of the peer-review committee on FGIN includes a list of 232 papers from FGIN published in peer-reviewed journals during its eight-year existence. There is evidence for impressive training activities (18 PhDs, 56 postdocs). Moreover, 25 scientists have spent sabbatical leaves at the institute.

We do not believe that your statements on FGIN's future reflects the thinking of the university leadership in view of the scientific and monetary benefits the university has received from FGIN. In our view, your leading article damages the scientific image of an institute that includes highly respected neuroscientists.

Giorgio Bernardi (President, Italian Society of Neuroscience); **Giovani Biggio**

(Chairman, Department of Experimental Biology, University of Cagliari); **Vittorio**

Ersamer (Professor Emeritus of Pharmacology, University of Rome); **Walter**

Fratta (Chairman, Department of Neuroscience, University of Cagliari); **Gian**

Luigi Gessa (Professor of Neuropsychopharmacology, University of Cagliari); **Paolo Mantegazza** (Rector,

University of Milan); **Flavio Moroni**

(Professor of Pharmacology, University of Florence); **Giancarlo Pepeu** (Chairman,

Department of Pharmacology, University of Florence)

■ In August, a spokesperson for Georgetown University said that the future of the Fidia-Georgetown Institute for Neuroscience was "questionable" if, as was considered likely, Fidia failed to honour its contract to help pay for its promised \$30 million share of the cost of new premises for the institute. The spokesperson says that the outlook has improved since then, and that "a new version of the institute may be constituted" when the financial reorganization of Fidia in Italy is complete. — Editor, *Nature*. □

No need to write

SIR — Hermann Bondi (*Nature* **365**, 484; 1993) deduces by a clear and simple argument that "the human mind is singularly liable to be mistaken on religious issues", and observes that "the variety of religions is a calamitously divisive force in human affairs". To anybody brought up in Northern Ireland, this conclusion and observation have been so self-evident that we had never (previously) realized that it merited a letter to *Nature*. *Res ipsa loquitur*.

Mark Cantley

*131 rue Verbiot,
B-1030 Brussels, Belgium*

Clash of the titans of space policy

Joan Johnson-Freese and George Moore

What does the future hold for space research? Wrangles between groups of people with different interests must be resolved if a prescription is to be found.

MUCH has been written about the flawed 92.5-inch primary mirror of the Hubble Space Telescope, most of the press reports concluding that the project has failed as a result. Space scientists, however, although certainly disappointed that the mirror is imperfect, believe differently — "failure" is not the word most of them would choose.

Part of the discrepancy in judging the success or failure of a space project stems from the different criteria used to determine 'success' by the interest groups involved: scientists, engineers, programme managers, politicians and the public. Engineers tend to declare a mission successful when the hardware works; scientists withhold judgment until the data returned can be evaluated; whereas programme managers, and ultimately the general public and subsequently those responsible for financial support, are mainly interested in high-profile projects, ignoring the rest. These different attitudes stem from huge differences in philosophies and priorities. Engineers define 'success' in terms of performance; researchers in terms of knowledge; and, in the United States, at least, managers and the public are looking for how well the project adds to what has been the only consistent goal of US space policy — global leadership in technology. These differences are important because they result in scientific and economic inefficiencies, providing a tenuous basis for continued political support. The chasm is particularly apparent in space projects because of their inherently dramatic nature.

US space programmes have suffered several widely publicized problems recently, the latest of which is the lost Mars Observer satellite, resulting in considerable general disillusionment. Because the public is primarily interested in the symbolic value of space, and recent events have tarnished this symbol rather than shining it, many space specialists are concerned about the falling away of public support. The economics of space flight will exacerbate these problems unless something is done.

Public perceptions

Those administratively 'in charge' of a space project are primarily concerned with 'pleasing' those to whom they are immediately responsible. In the case of NASA, this involves congressional authorization and appropriations commit-

tees, which determine future support. But 'space', both as an entity and as a field of activity, offers relatively little to members of Congress in terms of return on their 'support' investment. Beyond the jobs involved in keeping both the space shuttle and the communications-satellite industry operational, there are relatively few 'space-related' jobs that translate into constituent support. In the end it is usually the public's view of space activity, as determined by how well the hardware seems to work and therefore how well US technological expertise is exhibited, that becomes important in terms of how much one project will generate support for future projects. Public disillusionment with the space programme is reflected in funding decisions made by Congress, where space directly competes with social projects.

The high cost-per-pound to orbit has kept the 'commercialization' of space from becoming a reality. The importance of this factor cannot be overstated. A simple cost evaluation accomplished by taking the cost of one launch (which can range from \$500 million to \$1 billion, depending on the vehicle used), and dividing by 50,000 lb (the weight that can be placed into low-Earth orbit), yields a widely varying figure. Not all launch vehicles can be used for all missions; even small 'inexpensive' vehicles like Pegasus have a very high cost-per-pound ratio. Manned vehicles, of course, cost the most because of the system redundancy and other safety requirements that come into play. Manned vehicles should, therefore, never be used simply as space hardware launchers as they yield enormous cost disadvantages to no advantage. The bottom line, though, is that high launch costs make every launch a critical one and limit the number of opportunities available to researchers. High launch costs have been prohibitive to commercial ventures other than communications satellites, where this overhead can be passed along to the large consumer market.

Therefore, the symbolic value of space, which first motivated civil space activity in the 1950s and military space activity during the Cold War, has long since become outdated. To many, this symbolism is insufficient as the primary *raison d'être* for our interest in space, but it still remains as the main consideration for decisions on support of programmes. When the number of problems outweighs the degree of

technological expertise that can be showcased, the very foundations of the entire space programme begin to shake.

It is ironic that the space shuttle is considered one of NASA's shining accomplishments, with more than 50 successful launches and 'only' one failure, when it clearly is a failure in meeting its original goal of affordable, easy access to orbit. The shuttle can be considered a success only if judged on its symbolic value, and even then only after an inordinate amount of forgiveness for that one 'glitch'. The problem with space as a symbol is that space activity is inherently a high-cost, high-risk activity. Since Challenger, it has become increasingly difficult to argue that the symbolic value of what has been achieved, in contrast to the well-publicized failures, has been worth the money spent.

Scientists

Space scientists in the United States today face the sombre reality that in their probable 30-year career they may have the opportunity to fly only two or three space missions. Therefore, most must be extremely careful in choosing projects, as they may be committing themselves to a decade or more of work and a professional reputation that will flow from that project. Those who are successful and thus able to pick and choose their projects must also be concerned with their reputation. Many will not even consider an 'announcement opportunity' (a government agency offer of support for development of a specific technology or engineering project) unless they feel that the resulting scientific advance is very significant. If a scientist decides to propose an experiment for a space-science mission, the generally acknowledged first step is to define the scientific goals. Once these are determined, those specialists most competent to accomplish these goals are identified and assembled by the principal investigator. If this team is international, all the better as additional money as well as expertise results. Once the team is identified, and if the experiment or project is funded, the interaction between these scientists and the engineers responsible for building the hardware often becomes compartmentalized, even subtly adversarial.

Many scientific researchers feel that engineers, particularly project managers, would prefer to receive a list of require-

ments and have little to do with the researchers themselves. Researchers see a project as a dynamic effort requiring long-term commitment, whereas they believe that engineers prefer to remain detached in case it becomes necessary to sacrifice parts of the project for reasons of economy. The thought that politicians and administrators may be considering cutting off fully productive satellites such as the Extreme Ultra-Violet Explorer is repulsive to space researchers, yet economics and a lack of public involvement may indeed dictate just such an action.

Scientists want to be involved with the project design and hardware throughout the process, feeling that their participation will result in a better, and perhaps cheaper, end product. As an admittedly oversimplified example, scientists explain that if they simply give engineers a requirement for an aluminum box, the result may be that engineers spend an inordinate and unnecessary amount of money squaring the corners, assuming this to be necessary simply because a 'box' was requested. In reality, though, having a rounded-corner container may make no difference whatsoever to the scientific outcome.

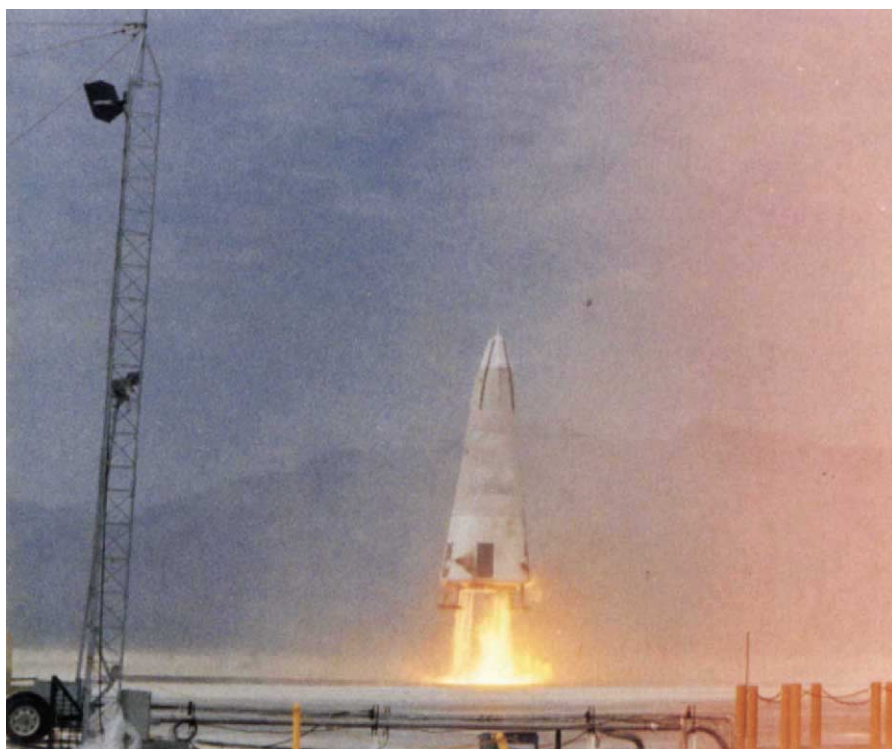
At the political level, cutting money from space science budgets is a relatively 'safe' decision unlike, for example, cutting a social programme or a programme with strong support from a congressman's constituents. It is even 'safer' to cut a non-manned space project, which appeals less to the public than the manned programme (except for fleeting periods of extraordinary interest, such as a viewing of Halley's comet or of a deep space planetary fly-by).

Many space scientists (although not life scientists and perhaps materials scientists) believe that space exploration can be carried out with less expense and more robustly through the use of robotics. It is also true, though, that they recognize the direct correlation between overall space funding, much of which is for manned space flight, and funding for space science. The general rule has been and remains that space science receives 20 per cent of NASA's budget.

A functional definition of when man is required in space exploration has been given by Edward Teller. In near-Earth operations, robotics can be precisely directed and controlled in real time or very near real time. But in distant planetary explorations where a robot might discover a significant sample or critical data, the time required to report to Earth, allow a man to make the appropriate decision on how to proceed and transmit directions back to the robot, means that man is required to be on or near the site of exploration.

Engineers

When a space science project is launched, it is the engineers who take the responsi-



Lost hope for the future? — The cancelled DC-X experimental technology demonstrator space vehicle, which would have provided cheaper flights.

bility for the success or failure of the project in that they feel the 'heat' if something goes wrong. Conversely, few individual engineers receive recognition if the project works as planned. Because it is assumed that the hardware will operate successfully, it is the scientists who obtain the glory when spectacular pictures are returned of Saturn's rings, Halley's comet, or whatever. Hence many engineers approach the project from a perspective of minimizing their risk, which is very different from that of scientists, who seek to maximize the scientific returns. Many engineers maintain that scientists are more inclined to take risks because they do not share the same responsibility for a mission's success.

In the former Soviet Union, for example, it was the former director of IKI, Roald Sagdeev, who was in the spotlight during the successes of the Vega missions, with the engineers in the background. With the 'failures' of Phobos I and II, however, it was the engineers who were blamed. Ironically, in this case, it was the scientists who insisted on hurrying the spacecraft development process, over the concerns of the engineers involved, to make the earliest possible launch window. This left the engineers with far less time for ground testing than they wanted, with what some would say were predictable results. At least partly because of that experience, what was to have been the Mars 92 mission has now become Mars 94 — engineers insisted on more time to test the hardware.

Engineers want to freeze a project's

design as quickly as possible so that they can proceed with trying to meet the requirements with the minimum of risk. Hardware reliability is clearly their priority, sometimes at the expense of the payload. When interplanetary trajectories are involved, for example, many course corrections are necessary, increasing the fuel requirements for a spacecraft, and thereby increasing the spacecraft weight. The power supply for the spacecraft and instrumentation is another critical factor of the engineering success of the mission, therefore engineers want to place as many batteries on board as possible, again adding weight. The durability of the craft itself is a 'risk' consideration, which can prompt engineers to strengthen the materials to be used — yet more weight. All this is weight given to the structure in preference to the instrumentation payload.

When scientists work with engineers to develop hardware (and software) for ground-based telescopes, they can experiment to advance the instrumentation through development. The long time between space-based missions, however, leaves no opportunity for such an incremental development process. This often means that scientists are locked into technological decisions sooner than they would like, so that the engineers can begin drawing their blueprints.

In the United States, the political 'cost' of failure, or even perceived failure, has become such that NASA is seen by many scientists and analysts as no longer interested in new technology but wanting to

stick with off-the-shelf, proven hardware. To a large extent, this is seen to be a function of NASA changing, perhaps being politically pushed, from an organization that was performance-driven to one that is cost-driven. The problem is further exacerbated by a bureaucracy whose culture demands that it thinks big, to perpetuate the organization and its image from the glory days of the 1960s. The reality, however, is that scientists and engineers are rarely allowed to turn dreams into hardware any more. With so few missions, there is little opportunity to establish a track-record of success. Every effort 'counts' in an over-magnified sense, demanding success.

Conflict

Every flight, every mission and every experiment is seen as a once-in-a-lifetime experience, or nearly that, to space scientists. Because of this frustration with the lack of opportunities, when one does arise, scientists want to take advantage of it. They want to use the latest technology in a world of fast-paced advancements. From the engineers' perspective, this means that scientists tend to change their minds frequently as to just what it is they 'want' or 'need'. The counter-response is that defining the whole goal is impossible — scientists want as much as they can get.

Because they know that risk will be minimized and consequently perhaps returns threatened, scientists initially tend to be over-optimistic in stating what they want to do on a mission. They have to be careful, however, not to be too zealous, or their credibility will suffer, making the entire process a kind of game where efficiency and returns can suffer because of anticipated technical and scientific trade-offs. On the other hand, scientists and engineers also build on each others' achievements and capabilities. With space science missions becoming smaller to avoid catastrophic losses of billion-dollar missions, scientists have sometimes been pessimistic about how much can be accomplished under new, more conservative guidelines. They have also, however, sometimes been pleasantly surprised at what engineers can accomplish within these guidelines. Challenges such as these have spurred on the engineers, which in turn have inspired scientists to stretch experimental goals.

Both scientists and engineers complain that NASA managers can be too conservative (which is at least in part because managers know that scientists will try to expand on a spacecraft's capabilities as it is being built). To take an example, NASA recently sent out an opportunity announcement concerning the small explorer (SMEX) programme. SMEX are to be small, relatively inexpensive experiments launched using a Pegasus rocket, within a three-and-a-half year time frame,

with an approximately one-year spacecraft lifetime. They are part of the small, quick and cheap philosophy that NASA administrator Daniel Goldin has tried to push, but which for the most part has not percolated down through the bureaucracy. To keep scientists conservative in their goals, NASA's announcement was very specific in terms of which Pegasus model could be used, which subsequently limited the payload weight. In some cases, it was clear that the experiments proposed could have been done better and cheaper on a newer Pegasus model, but NASA did not leave the flexibility to allow a change.

The dilemma, too often it seems, comes back to money. Supporters of the civilian space programme, particularly those concerned with scientific research, are fighting to defend goals and objectives which, to many politicians pressured by constituents facing short-term economic problems, are superfluous at best. Despite a very large US military space budget, there is little direct financial spillover benefit to civilian space efforts. In national-security space programmes, cost is seldom a significant factor — the technology in use may be far from cost-effective and therefore not useful for a civilian project. Although some national laboratories are concerned with space-science related research, for example X-ray astronomy research at Los Alamos, most space-science research is conducted by academics not supported by the Department of Defense. There is some technology spillover, such as infrared arrays originally developed for missile tracking being modified and improved for infrared astronomy in studying the birth of stars and evolution of galaxies, but spacecraft and launch expenses still require the lion's share of available funding. As spacecraft size and weight limitations are often determined by launch limitations, this brings the argument full circle to the issue of how to reduce the cost-per-pound to orbit.

Where next?

With increasingly stringent budget conditions becoming the norm in the United States, the present situation of conflicting goals for space activities cannot continue for much longer. Scientists, viewed by many as being given extraordinary amounts of money to play in their heavenly sandboxes, will be increasingly questioned about their goals and purposes. Engineers who need scientists to provide them with missions (manned or unmanned) have to appreciate that their future will be realized only through a symbiotic relationship with scientists. Managers need to appreciate the problems of each group. It must also be acknowledged that any goal has two equally critical aspects: a properly functioning system or spacecraft; and the acquisition of valuable scientific

data.

This situation is not inherently adversarial, but has evolved because of economic circumstances. Leadership — at the national, bureaucratic and project level — must emphasize the value of attaining the common goal, however defined in terms of science or of exploration, and somehow equally recognize the contributions made by all involved. That would go a long way towards dealing with the animosity among these groups and would promote a more cooperative spirit.

The more complex issue is that of defining the overall goal of a space programme. Space activity in the United States was prompted by the desire to exhibit the country's technological leadership. In future, it is in danger of being dictated by bureaucratic parochialism, political stalemates and by viewing space activity as expendable or at least protractable, rather than having clear goals or technological use.

Although the situation is most pronounced in the United States, much the same problems exist elsewhere. Unless determined efforts are made to set goals for space activity in the next century, and to develop the required technology in a timely and cost-efficient manner — even if that violates norms of the past — the clash of the titans may be a battle to the death for all space programmes.

The prescription for the future US space programme will require leadership at the highest levels, within and among NASA, Congress and the President's office. So far, that leadership has not been forthcoming. Over the past 15 years, study after study has been carried out to "set the course" of the US space programme, all advocating the development of a lower-cost space transport vehicle as a top priority. For the most part the United States, and the rest of the world, is still using technology from the 1960s and 1970s. In the latest showing of intransigence from the past, the US Congress in September 1993 refused to support further funding of the McDonnell-Douglas DC-X, single-stage-to-orbit, technology demonstrator programme, which had shown substantial potential for addressing the cost-per-pound-to-orbit issue head-on. Although the technology required to make the vehicle operational still has a long way to go, the notion that a spacecraft could be built without an accompanying gargantuan infrastructure, stacks of diagrams weighing the same as the vehicle, and an army to accomplish its launch, is refreshing and uniquely appealing to supporters. Unfortunately, the paradigm of the past seems to be prevailing, but it is by no means too late to change. □

Joan Johnson-Freese and George Moore are at the Air War College, Maxwell Air Force Base, Alabama 36112, USA.

Social science and the new world order

The Fourth Indira Gandhi symposium last week may not have redefined the good society with precision, but it did succeed in showing that the path into the future is beset with contradictions to be resolved.

New Delhi. Mrs Elena Bonner, widow of Andrei Sakharov, raised one of the few discordant notes at the Fourth Indira Gandhi symposium in New Delhi last week by pleading that there should be no reference to Mr Mikhail Gorbachev in any statement put out by the meeting. Her case was that he had justly become deeply unpopular in what used to be the Soviet Union, and that he should not be accorded credit, but rather contumely, for recent historical events. But the plea was quickly denied by a compatriot, who said she disagreed with every word of it. Mrs Bonner herself seemed unaware that she was asking that history should be rewritten, as it was in Stalin's time.

Mrs Bonner's indiscretion was not the only spectacular instance of stereotype reversal at this meeting of the great, the good and the occasionally rapscallion. Thus Mr Robert McNamara, President John F. Kennedy's Secretary of Defense, who soldiered on through much of the Vietnam War, startled some who had not followed his succeeding career as president of the World Bank with an urgent statement of the case for a nuclear-free world, punctuated by liberal table-banging. In a more minor key, Sir John Kendrew, long a member of the British science establishment, asked that more attention should be paid to the social sciences and to what they have to say about the improvement of society; natural scientists in Britain have hitherto been conspicuous for their scorn of the softer sciences.

The symposium itself, at least in its immediate effect, may have been more instructive for its participants than for those who will in due course read the record of what was said in the three and a half days. The topic, "Redefining the Good Society", chosen by the Indira Gandhi Foundation on the grounds that the world has changed with the ending of the Cold War and the break-up of the Soviet Union, proved to be an opportunity for the participants to reflect generally on the future of mankind. But Dr Kenneth Kaunda, president of Zambia until 1992 and chairman of one of the seven sessions, was so moved by his 40-minute account of his past teenage oppression by the British that he was in tears half-way through. The Prime Minister of India, Mr P. V. Narasimha Rao, may have been more canny than wistful when, in opening the proceedings, he wondered whether "redefining" should not have been replaced by "rediscovering".

What follows is one participant's opinion about which of the several issues raised are likely to endure.

The problem of national (and thus international) security, is one of these, and McNamara's view is the most forceful. Indeed, he has a vision of what the 'new world order' should be: frontiers never changed by force, redress for minorities and ethnic groups deprived (by their governments) of their rights, conflict resolution without the unilateral intervention of major powers, more technical and financial assistance for developing countries and sustainable development for all. McNamara holds that the end of the Cold War does not necessarily mean a return to nineteenth-century power politics; collective security is an alternative.

Controlling strategic weapons is one means to that end. McNamara demanded support for an extension of the arms control agreements negotiated in the past two decades. His audience nodded approval. He told a chilling tale of a series of meetings with Russian counterparts at which it has become plain "how close the world was to nuclear catastrophe" during the Cuban missile crisis of 1962; contrary to US intelligence, 162 Soviet warheads had already been transferred to Cuba and would have been used to repel an expected invasion of the island, with the certainty of subsequent escalation. So, speaking from 30-year-old experience, McNamara wants to work towards a nuclear-free world, to which end he wants the UN Security Council to have powers to coerce unwilling governments to sign the Non-Proliferation Treaty. He mentioned North Korea, to general approval; he did not mention India, thus avoiding certain ructions. But the argument will break out at some stage.

Development is properly another insistent theme, divisible four ways by the orthogonal discontinuous variables of population growth and the rest along one axis and micro- and macro-economic considerations along the other. Everybody agrees that rapid population growth is a curse at both levels; the Indian prime minister, indeed, said as much. Many Indian participants insisted that India's economic growth is regularly cancelled out by population growth. For the past decade, economic growth has averaged 6 per cent compared with population growth of 2.4 per cent a year; the trouble is that inflation accounts for much of the difference and the enlargement of the middle classes may have eaten up what is left, leaving nothing to benefit the poor.

So the hunt is on for techniques other than general and substantial enrichment to engineer the classical demographic transi-

tion from a high to a low birth rate, which in India has fallen by a third in 20 years (to 30.2 per thousand). Kerala, the southwestern Indian state with just under 30 million people, is everybody's shining example. Infant mortality (17 per thousand) is less than a quarter of that in the rest of India, the average age of women at marriage is 22 years (compared with 18), the female literacy rate is 87 per cent (twice the national average), the expectation of life at birth is nearly 72 years and, despite an average annual income (in rupees) of just over US\$100 a head, the net replacement rate is within a whisker of the magic figure of 1.0.

This example shows that the technology of contraception may be a necessary but is not a sufficient catalyst of the classical demographic transition. That is why the symposium united behind the social science solutions — better public health, sending children (especially girls) to school and what Marjorie Thompson, former chair of Britain's Campaign for Nuclear Disarmament, called "the empowerment of women". Can the example be copied elsewhere?

What part does science play in all this? With reservations, the symposium seems to have demonstrated that the developing world still regards science and its intelligent application as its salvation. Has it not, after all, shown that the world can be fed? — part of the powerful case put by Dr Bernard Lown, the Boston cardiologist who helped to win the Nobel Peace prize in 1985 for International Physicians for the Prevention of Nuclear War.

But as on development generally, there is a gulf between the micro- and macro-techno-enthusiasts. The former (indistinguishable from the German Greens, and represented last week by Dr Ekkehart Krippendorff, who is one of them) welcome higher yielding strains of crop plants but insist that not everybody can be as prosperous as "the Americans", who must therefore become poorer in the cause of equity. Others, like Yash Pal, the physicist who borrowed time on other people's satellite transponders to bring educational television to Indian villages, regret the present influence of cable television on India's youth.

Certainly nobody last week held that letting technology rip would solve as many problems as it would create. That is the crux of the case for the social sciences. But last week's representatives of those arts were even more given to generalities than is their everyday habit. Perhaps the next meeting should be in Kerala.

John Maddox

Methuselah among nematodes

Linda Partridge and Paul H. Harvey

MYTH and literature have given human immortality mixed reviews. There is, nonetheless, fairly general agreement that intimations of mortality, in the form of ageing or senescence, are regrettable and should be postponed as long as possible.

On page 461 of this issue¹, Kenyon and co-workers report a mutation of the nematode worm *Caenorhabditis elegans* that more than doubles its healthy and fertile adult lifespan. At first sight this vermiform medical marvel, the equivalent of a healthy 200-year-old human, produces something of a headache for evolutionary biologists because evolutionary theories of ageing strongly predict that no single mutation should be able to have such a dramatic effect^{2,3}. In essence, this is because death and infertility as a result of ageing are expected to be a consequence of many processes, affected by many genes; fixing one of them will therefore simply leave others to intervene very little later than before. So how has the worm done it?

The reason is probably that the mutant gene controls the activities of many downstream genes, because it is displaying part of an alternative life-history pattern⁴ to the one usually seen in the laboratory, and one that is normal and adaptive under some of the circumstances that the worm encounters in nature. The ordinary wild-type worm can take two different routes through the pre-adult period of growth. Under normal circumstances, and those usually prevailing in the laboratory, the hatchling progresses smartly through a series of larval instars to the reproductive adult stage. In contrast, if times are hard because of crowding or food shortage, a decision is taken in the first larval instar to form a dauer, a developmentally arrested version of the third instar, which is small, thin, sexually immature and very long-lived. The worm can presumably survive bad times in this way, and develop through to adulthood if conditions improve.

The mutations that so dramatically extend adult lifespan are in one of the genes (*daf-2*) in the developmental pathway leading to dauer formation, and they increase lifespan in the absence of the environmental cues that usually trigger it. A twist is that the mutants are temperature-sensitive, and only display the mutant dauer phenotype at a higher, non-permissive, temperature than that at which they behave like wild type. Kenyon *et al.* could therefore allow the mutant worms to develop through the first larval instar at the lower, permissive temperature, so that they did not enter the dauer stage.

Thereafter, whether they were compared at the permissive or the non-permissive temperature, the *daf-2* mutants lived longer than wild types. *daf-2* must therefore have other effects after the dauer stage, and the fact that the mutant adults are longer lived even at the permissive temperature is perhaps because there is still some deficit in *daf-2* function. The mutants were slightly less fecund than wild type under laboratory conditions. But, by

offspring that enter the breeding population. By contrast, a worm that is forced through the dauer route is likely to encounter less than optimal breeding conditions as an adult, and would gain by building as durable an adult soma as possible, to allow offspring to be produced over a long period with the chance that conditions will improve. It would be interesting to know how these two types of adult fare under conditions of food shortage, and whether the dauer-route adult has a longer pre-adult growth period than wild type in the absence of dauer production.

The consequences of the mutations in

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Wriggling into the record books — *Caenorhabditis elegans*, the species in which Kenyon and colleagues identified lifespan-enhancing mutations, and thereby provide evolutionary biologists with food for thought. (Sinclair Stammers/Science Photo Library.)

laser ablation of the precursors of the germ cells and the somatic gonad, the authors showed that the longevity advantage of *daf-2* was unaffected, so that any costs of making gonads or germ cells after the time of laser-treatment could not explain the shorter lifespan of the wild-type worms.

The dauer route to adulthood seems to be a mechanism for sitting out hard times. It is therefore perhaps not surprising that this route produces a different kind of adult, aside from any direct after-effects of having entered the dauer stage itself. Theories of the evolution of life histories have shown that prevailing demographic conditions affect the value of offspring produced at different times⁵. If the population is expanding, offspring produced early in life are more valuable to their parents than those produced later; that's because the early offspring will themselves start breeding sooner and will contribute more to the expanding numbers by more rapid compound interest. If numbers are declining, then offspring produced later are at a premium because they contribute genes to a smaller pool.

A larval worm that encounters good conditions for growth is likely to encounter good conditions as an adult, and would therefore benefit from rapid production of

daf-2 can therefore be regarded as the equivalent of changes in many genes rather than just one, because *daf-2* controls the state of many downstream genes active at different stages in the life history. Normal senescent decline, of the kind seen within both the dauer and non-dauer route adults, is expected to be independent of the developmental switch between the routes. The hierarchy of genes controlling development functions in a precisely co-ordinated fashion that ensures each sequence is expressed in the right part of the body at the right time. These genes are developmental specialists — they themselves have evolved to their present sequences because they enabled the complex phenomena of embryology and growth to happen in a repeatable and orderly fashion. Whatever the genes that affect ageing, they are not expected to have evolved to their present form because they themselves brought it about, and 'a regulated mechanism of ageing' is not expected to exist.

1. Kenyon, C., Chang, J., Gensch, E., Rudner, A. & Tabtiang, R. *Nature* **366**, 461–464 (1993).
2. Charlesworth, B. *Evolution in Age-Structured Populations* (Cambridge University Press, 1980).
3. Partridge, L. & Barton, N. H. *Nature* **362**, 305–311 (1993).
4. Gross, M. R. *Nature* **313**, 47–48 (1985).
5. Cole, L. C. *Q. Rev. Biol.* **29**, 103–137 (1954).

Ageing happens because things go wrong as a result of damage and failure. Some of these problems are partially prevented or repaired, for instance by enzymes that mop up free radicals or replace damaged sequences of DNA, and the activities of these genes will therefore affect the rate of ageing. So too will genes that affect the rate at which damage is accumulated, for instance through reproductive or metabolic activity. Genes involved in growth and development will also be important, because they can affect the specifications of the adult soma, and hence its ability to resist damage. Most

genes are therefore likely to influence the rate of ageing — however they would be regarded as having some other function, would not be recognized as 'senescence genes' as such and their individual effects on the rate of ageing will be small. □

Linda Partridge is in the Institute of Cell, Animal and Population Biology, Division of Biological Sciences, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, UK. Paul H. Harvey is in the Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

EVOLUTIONARY PHYSIOLOGY

Quantitative design of life

Jared M. Diamond

EVOLUTIONARY physiology is a young science with a lot to offer, not least in providing a unifying framework for the current surge of descriptive studies in cell and molecular biology. An example of what can be achieved is to be found in the journal *Oikos*¹, in the form of a report by Koteja and Weiner who have tackled the issue of the evolution of basal metabolic rate.

Measurements of basal metabolic rate (BMR, defined as the metabolic rate measured at rest in a non-digesting adult animal at environmental temperatures within the thermoneutral zone) are fundamental to physiology. Different species of the same body mass can vary greatly in their BMR, the most marked difference being the some-ten-times higher values characteristic of endotherms (such as birds and mammals) against those of similar-sized ectotherms (such as reptiles and fish). But there are also considerable differences in the BMRs of endotherms of the same size and ectotherms of the same size. Koteja and Weiner set themselves the task of trying to tease out the regularities that may underlie this variation.

Correlates

Much of the variation is clearly associated with taxonomy^{2,3}; for example, marsupials have generally lower BMRs than placental mammals of similar size. Several authors have also suggested that BMR may be related to life-history variables⁴⁻⁶; for example, that BMR may tend to be low in desert species or in mammals that eat ants. However, it has been difficult to disentangle these putative life-history correlates from the well established taxonomic correlates because taxonomically related species tend to be similar in their life histories.

Koteja and Weiner used two methods to cut this Gordian knot. First, they sought to minimize taxonomic variation

by confining their examination to a single mammalian group — the muroid rodents ('rats and mice' to the general public). This very successful stock has radiated to produce hundreds of species that are diverse in their life histories and ecology. Second, Koteja and Weiner recognized that values of life-history traits (such as diet, body size and litter size) do not vary independently of each other but tend to fall into clusters: for example, large mammals tend to have litters of few pups. Natural selection acts not on individual traits but on individual animals and on their clusters of traits. Hence Koteja and Weiner sought to identify such clusters and to detect correlations of BMR with the clusters rather than with individual traits.

For each of 90 muroid rodent species, they tabulated values of four variables or sets of variables. One was of course BMR. The second was body mass, the main determinant of BMR, whose influence must be factored out before any other effects can be detected. The third was taxonomy: the assignment of each species to one of the three recognized muroid families (Muridae, Cricetidae, and Arvicolidae). The remaining set of variables consisted of values of four life-history traits: diet (ranging from plant foods of low nutritional value to animal foods), climate (ranging from boreal to tropical), vertical foraging stratum (ranging from arboreal to underground) and habitat (ranging from desert to swamp).

Factor analysis and cluster analysis showed that numbers defining each species' position along these four life-history axes fell into three clusters, corresponding to the ecomorphological strategies intuitively associated with 'voles', 'mice' and 'hamsters'. One cluster (voles) eats a low-quality diet and lives in cold climates; a second cluster (mice) eats a high-quality diet and lives in warm, dense, wet habi-

tats; and a third cluster (hamsters) also eats a high-quality diet but lives in warm, open, dry habitats. It should be emphasized that these clusters are defined strictly by life-history variables and not at all by taxonomy. In fact, with just one exception, all three muroid families recognized by taxonomists are represented among all three life-history strategies recognized by cluster analysis.

Residuals

So, taxonomy and life history are sufficiently independent of each other among muroid rodents to enable their effects on BMR to be teased apart, by analysing residuals from regressions of BMR on body mass. Among species with the same cluster of life-history traits, species from the taxonomically defined family Arvicolidae prove to have higher BMRs than species from the other two families. Among species from the same family, those of the hamster and mouse clusters prove to have the lowest and highest BMR respectively. These correlations of BMR with life-history strategy are also apparent within single genera containing representatives of different life-history strategies: for example, the archetypal rodent genus *Rattus*, which includes species with the mouse strategy as well as species with the hamster strategy. Thus, both taxonomy and life-history variables significantly influence BMR.

What might be the proximate mechanisms underlying these differences in BMR? Accumulating evidence supports the hypothesis that animals with unusually high BMRs for their body mass have disproportionately large masses of some relatively small organs with very high mass-specific metabolic rates⁷. These organs include notably the heart and kidney, and also the liver, lung and small intestine. The correlations between these organ masses and BMR emerge from intraspecific as well as interspecific comparisons. For example, tropical bird and mammal species, which have lower BMRs than temperate species of similar size, have relatively small masses of the heart, kidneys and liver⁸. Among 22 bird species (mostly individuals collected in The Netherlands), relatively large hearts and kidneys proved to be the best predictors of high positive residuals from BMR-versus-body-mass regressions⁷. This covariation between BMR and organ masses is both genetic and reversibly phenotypic. For example, kidney mass relative to body mass has a very high coefficient of genetic determination in inbred mouse strains that covary in BMR⁹. At the same time, relative masses of the kidney and other energetically expensive organs vary reversibly with food intake (to which BMR is adjusted) in kestrels maintained on high- or low-calorie diets¹⁰ and in mice whose food intake is manipulated

by varying the energy demands of thermoregulation¹¹.

These observations immediately raise the question of why an animal should incur the expense of large, energy-consuming organs. Here we pass from the realm of proximate to ultimate causation. High BMRs are linked to high daily energy budgets in wild animals¹², a link which arises from the fact that high energy budgets require big metabolic machinery: big intestines to absorb the nutrients, big livers to process them, big hearts to pump them around, big lungs to perform the gas exchange necessary for their oxidation, and big kidneys to excrete the resulting wastes. BMRs and daily energy budgets thus become subject to evolutionary cost/benefit trade-offs. All other things being equal, high energy budgets yield benefits such as behavioural dominance, high growth rates and high reproductive rates. But high budgets also incur offsetting costs: animals with energetically expensive metabolic machinery cannot operate economically in environments or lifestyles of low productivity. Thus, organ masses and energy budgets become co-optimized through natural selection.

Koteja and Weiner's paper exemplifies the approach of evolutionary physiology, which attempts to explain quantitative values of physiological parameters as outcomes of natural selection. In designing machines, engineers routinely take account of expected loads on machine

parts when specifying quantitative values of the capacities designed into those parts¹³. Natural selection in effect plays the same role in matching biological capacities to their natural loads. Physiology, the science of biological function, could have made major contributions to our quantitative understanding of natural selection by evaluating those matches. But most research in cell and molecular biology is descriptive, in the sense that it stops short of asking what ultimate selective factors explain why this animal ended up with these particular capacities, rather than with higher or lower capacities. Simultaneous examination of animals' metabolic rates, life-history variables, organ masses and capacities at cell and molecular levels may enable us to interpret biological design quantitatively. □

Jared M. Diamond is in the Department of Physiology, University of California Medical School, Los Angeles, California 90024, USA.

1. Koteja, P. & Weiner, J. *Oikos* **66**, 505–514 (1993).
2. Elgar, M. & Harvey, P. *Funct. Ecol.* **1**, 25–36 (1987).
3. Harvey, P. *et al.* *Am. Nat.* **137**, 556–566 (1987).
4. McNab, B. *Ecol. Monogr.* **56**, 1–19 (1986).
5. Dawson, W. R. J. *Mammal* **36**, 543–553 (1955).
6. Haim, A. S. *Afr. J. Sci.* **83**, 639–642 (1987).
7. Daan, S. *et al.* *Am. J. Physiol.* **259**, R333–R340 (1990).
8. Rensch, I. & Rensch, B. *Zool. Anz.* **156**, 106–124 (1956).
9. Schlager, G. J. *Hered.* **59**, 171–174 (1968).
10. Daan, S. *et al.* *J. Biol. Rhythms* **4**, 267–283 (1989).
11. Toloza, E. *et al.* *Am. J. Physiol.* **261**, G608–G620 (1991).
12. Koteja, P. *Funct. Ecol.* **5**, 56–64 (1991).
13. Alexander, R. *Sci. Prog., Oxford* **67**, 109–130 (1981).

COSMOLOGY

Slaying the sacred cows

James Binney

ELSEWHERE in this issue (*Nature* **366**, 429–433; 1993), S. D. M. White and colleagues argue that clusters of galaxies, where dark matter was first detected, do not contain as much dark matter as is predicted by current cosmological models. They are led to this conclusion by comparing the amount of ordinary (baryonic) matter in the Coma cluster with the mean density of baryons in the Universe that is predicted by the standard theory of the synthesis, three minutes after the Big Bang, of the light elements (deuterium, helium and lithium). They argue that either our understanding of the origin of the light elements is in error, or we must come to terms with a Universe that has a lower than expected density.

From optical and X-ray observations, it is quite easy to place a lower limit on the baryonic mass, M_b , of the Coma cluster. First, one counts the total light radiated by the cluster's galaxies and multiplies by the mass-to-light ratio known to be characteristic of such galaxies from detailed studies of the internal dynamics of nearby galax-

ies. Next, one infers the mass of the hot gas that fills intergalactic space in the cluster from the cluster's observed brightness profile in X-rays. The latter has recently been remeasured by the Rosat satellite, and shows that within a fiducial radius of 1.4 megaparsecs of the cluster centre, roughly five-sixths of the baryons are in diffuse gas rather than galaxies.

Determining the total gravitating mass M_{tot} within 1.4 Mpc of the cluster centre is harder. Independent estimates come from models of the motions of the cluster galaxies and the assumed hydrostatic equilibrium of the hot gas. The problem with using the cluster gas is that its temperature at radii of interest is rather uncertain. The problem with using the cluster galaxies is that we do not know how elongated their orbits through the cluster are, and the more elongated a typical orbit is, the smaller is the predicted confining mass. Also, we know that the cluster is significantly nonspherical because its projection on the sky has an axis ratio of 2:1, but we do not know its

three-dimensional shape. This leads to additional uncertainty in its dynamical mass.

White *et al.* use N -body simulations of gravitational clustering to resolve these uncertainties: they 'observe' clusters formed in computer simulations and compare their true masses with those one would infer from 'observations' of them. They conclude that the ratio of the baryonic to the total matter content of Coma's interior out to 1.4 Mpc, M_b/M_{tot} , is greater than about 0.06.

Standard nucleosynthesis theory is able to predict the average baryonic density ρ_b of the present Universe because the mix of light elements produced as the cosmic temperature T falls through about 0.2 MeV is sensitive to the baryon density at that time. This density is generally expressed as a fraction of the mean density ρ_c required to halt the cosmic expansion. One finds $\Omega_b \equiv \rho_b/\rho_c \approx 0.012h^{-2}$, where the Hubble constant $H = 100h \text{ km s}^{-1} \text{ Mpc}^{-1}$.

Cosmologists have long inclined to the belief that the mean density of the Universe ρ is the critical value ρ_c . If so, the baryonic fraction of the Universe as a whole is smaller by a factor $0.012/0.06 \approx 1/5$ than that of the Coma cluster. This is not in itself worrying, because as clustering develops in the Universe, baryons tend to separate from nonbaryonic material; when a stream of the cosmic mix ploughs into another such stream, the baryons in each stream interact strongly and share their momentum, whereas the nonbaryonic material of each stream continues unperturbed on its way, as it interacts only weakly with other matter. The Solar System consists overwhelmingly of baryons for precisely this reason. But the larger the volume and thus the lower its density, the closer its baryon fraction should be to the cosmic value; the segregation of baryonic and nonbaryonic material on any scale cannot happen in less than the time it takes for an orbiting object such as a galaxy to cross the volume in question, and for a volume of mean density ρ this time is proportional to $\rho^{-1/2}$. White *et al.* argue in two ways that the Coma cluster is too large to have achieved a significant enhancement in its baryon content by the present epoch.

First, they treat the cluster as spherical and imagine that when each originally expanding shell of material around the cluster fell back into the cluster, its baryons were stripped out and remained in the cluster core, while its nonbaryonic material went on to highly elongated orbits. From this simple model they estimate that the baryon fraction within their fiducial radius should be a factor $Y \approx 1.4$ larger than the cosmic value, rather than the factor Y of 5 or more suggested by the observations. Their second argument involves simulations of cluster formation in

which nonbaryonic matter is represented by collisionless particles and the motion of baryons is followed by smooth-particle hydrodynamics. These simulations yield even smaller baryonic enhancements than the spherical model.

White *et al.* conclude that one or other of cosmology's sacred cows has to be slaughtered. Is the mean cosmic density ρ lower than ρ_c , giving $\Omega \leq 0.16h^{-1/2}$? Studies of large-scale streaming, which tend to imply that $\Omega \approx 1$, are probably compatible with such a value of ρ , but it is improbable on first principles. Although it implies that most of the matter in the Universe is of unknown nonbaryonic form, it conflicts with the standard argument for ρ being precisely equal to ρ_c : from the general theory of relativity we know that if ρ lies close to ρ_c today, in the past it must have lain extremely close, and in the future (assuming that $\rho < \rho_c$), it will become negligible compared with ρ_c . So if ρ is not equal to ρ_c , we happen to live at a most unusual epoch and one wonders why ρ was originally so close to a special value without being precisely equal to it.

Is standard nucleosynthesis theory at fault? If the Universe was inhomogeneous at the time of nucleosynthesis, the mean baryon density could be higher than standard theory predicts. It has been speculated that the required inhomogeneity was generated by a first-order quark-hadron phase transition, but such models are now considered implausible.

The fault more probably lies with our ability to estimate the expected baryon enhancement Y on the scale of clusters. There are two reasons to be cautious. First, Y must depend on how well developed clustering is on scales larger than those of clusters. The standard cold dark matter picture, within which White *et al.* work, underestimates this. Second, in this picture, small structures form first and ever larger structures form by the merging of smaller structures. This process can be accurately simulated only in so far as small-scale structure is genuinely eliminated at each stage in the merging hierarchy, as numerical techniques have limited spatial resolution, and cannot simulate the dynamics of systems in which a wide range of scales are active simultaneously. Unfortunately, the structures we are most familiar with — our Galaxy and galaxy clusters — do seem to involve a wide range of scales.

So although the argument of White *et al.* is interesting, it would be rash to abandon ideas that relate to times when the Universe was fairly homogeneous, and thus easily modelled, just because they conflict with simulations of a later, more theoretically challenging, epoch. □

James Binney is in the Department of Physics, University of Oxford, 1 Keble Road, Oxford OX1 3NP, UK.

Trail of the mystery bands

Theodore P. Snow

ON page 439 of this issue, Fulara and co-workers¹ describe the latest in a series of attempts to discover the particles responsible for the oldest unsolved problem in astronomical spectroscopy: the unidentified diffuse interstellar bands (DIBs). The first of the DIBs was noticed in 1921 (see ref. 2), and today well over a hundred of these broad, shallow interstellar absorption features have been listed^{3,4}. Speculation over their origin has oscillated between molecules and dust particles, with molecules being favoured by most current researchers.

Fulara *et al.* have found a number of near-coincidences in wavelength between observed DIBs and laboratory spectra of unsaturated hydrocarbon molecules deposited in neon matrices. Besides proposing a new candidate for the 'carrier' of the DIBs, their work is a good example of the new interest being shown by laboratory chemists in what has mainly been a concern of astronomers. It contributes to the renaissance of DIB research that began a few years ago as astronomical detector systems became more sophisticated, and was then fuelled by the discovery that large carbonaceous molecules may be common in the interstellar medium. These look particularly good candidates for the DIB carriers: the lack of clearly defined vibrational series or rotational fine structure tends to argue against small molecules, and the large combined absorption cross-section represented by the DIBs requires a carrier made of cosmically abundant elements. In an era when astronomers speak glibly of species such as polycyclic aromatic hydrocarbons (PAHs)⁵ or even fullerenes⁶ in the interstellar medium, large carbonaceous species are obvious suspects in the great DIB mystery^{7,8}.

In spite of the growing body of evidence for such molecules in the diffuse interstellar medium, little work has been done on how these large species are formed and dispersed. The only molecules known to exist in the clouds that produce the DIBs are diatomics such as H_2 , CO and the like, and the chemistry that produces these simple species certainly cannot also account for molecules such as PAHs or fullerenes. Condensation in outflows from carbon-rich red giant stars⁹ is actually thought to be a more likely source than *in situ* formation through gas-phase chemistry. The species suggested by Fulara *et al.* as the DIB carriers could be related to the carbon-chain molecules that are seen in dark clouds, and it has been proposed that carbon chains may grow through gas-phase reactions in these dense regions (P. Thaddeus, personal

communication). But it is still unclear how the molecules are then dispersed into the more rarefied medium where the diffuse bands arise.

Fulara *et al.* report laboratory studies of hydrocarbon species of the generic form C_nH_m , where n is in the range 6 to 12 for the species studied so far, and m is small, so that the species are unsaturated. The measurements were made in the solid phase; the molecules are formed as negative ions which then lose their excess electrons during deposition with neon into a matrix. Within the spectral region 600–1,150 nm, many absorption features were seen, several of which coincide (to within 0.5 nm) with observed DIBs. The wavelength matches are not precise, possibly because of shifts due to matrix effects, and there are many known DIBs that have no counterpart in the laboratory measurements. Furthermore, there are a number of features seen in the laboratory spectra for which no DIB is known; so Fulara *et al.* urge observers to seek DIBs at those wavelengths. It is not clear that Fulara and co-workers have found a specific carrier, but they may be on the trail of a class of molecules that is responsible for at least some of the DIBs.

Other equally tantalizing matches between laboratory and interstellar features do exist. Last year, for example, laboratory studies of two different species of positively ionized PAH molecules (again trapped in a solid matrix) produced results similar to those of Fulara and co-workers. Close coincidences were found, but other features were seen in the laboratory that did not correspond to known DIBs^{10,11}. And at least one other class of large molecule, the porphyrins, has been put forward as the DIB carrier, again on the basis of laboratory measurements of matrix-isolated molecules¹².

It is gratifying that the problem has gained such visibility in the chemistry community, and to see the synergism that

1. Fulara, J., Lessen, D., Freivogel, P. & Maier, J. P. *Nature* **366**, 439–441 (1993).
2. Heger, M. L. *Lick Obs. Bull.* **10**, 146–147 (1992).
3. Herbig, G. H. *Astrophys. J.* **196**, 129–160 (1975).
4. Herbig, G. H. & Leka, K. D. *Astrophys. J.* **382**, 193–203 (1991).
5. Puget, J. L. & Leger, A. A. *Rev. Astr. Astrophys.* **27**, 161–198 (1989).
6. Kroto, H. W., Heath, J. R., O'Brien, S. C., Curl, R. F. & Smalley, R. E. *Nature* **318**, 162–165 (1985).
7. Crawford, M. K., Tielens, A. G. G. M. & Allamandola, L. J. *Astrophys. J.* **293**, L45–L48 (1985).
8. Leger, A., d'Hendecourt, L., Verstraete, L. & Schmidt, W. *Astr. Astrophys.* **203**, 145–148 (1988).
9. Frenklach, M. & Feigelson, E. D. *Astrophys. J.* **341**, 372–384 (1989).
10. Salama, F. & Allamandola, L. J. *Astrophys. J.* **395**, 301–306 (1992).
11. Salama, F. & Allamandola, L. J. *Nature* **358**, 42–43 (1992).
12. Johnson, F. M. *Mem. Soc. R. Sci. Liège, Six. Ser. Vol. III*, 391–407 (1972).

now appears to be developing between laboratory chemistry groups and astronomers in pursuit of the mystery particles. Let us hope that the chemists and the astronomers continue to talk to each other, so that the many astronomical constraints are not overlooked by the chemists, and so that the context and implications of the laboratory work are fully understood by the astronomers.

But is progress being made? It is difficult to say, until more precise matches can be found. One hopes that it will soon

become possible to make laboratory measurements of candidate molecules and molecular ions in the gas phase, eliminating solid-state matrix effects. Despite the elusiveness of the DIB carriers, today one senses a new optimism that has been lacking in the past. Surely, it is tempting to think, we are on the right track, and it is only a matter of time until the definitive experiment is done.

Of course, the result may not be simple. We may find that several different species of molecules are involved, and that each is

responsible for one or at most a small number of the DIBs. If so, it will take time to unravel the entire problem. But when this is done, a wealth of new information on interstellar physics and chemistry will have been revealed, and a most intriguing astronomical mystery will have been solved. □

Theodore P. Snow is in the Center for Astrophysics and Space Astronomy, University of Colorado, Campus Box 389, Boulder, Colorado 80309, USA.

OBITUARY

Severo Ochoa (1905–93)

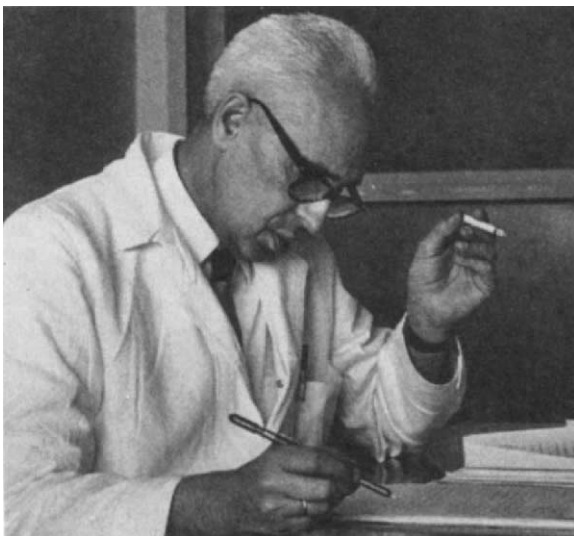
As one of the very few who stayed in the rapidly moving spotlight of biochemistry for five decades, Severo Ochoa bridged the evolution of the discipline from its origins in physiology to the heyday of enzymology and into the currents of molecular biology. With his death in Madrid on 1 November, Spain lost its most celebrated scientist and the international community an exemplar of science and statesmanship. The award of the Nobel Prize for Physiology or Medicine in 1959 made him the only other Spanish laureate in science since the neuroanatomist, Ramon y Cajal, who was so honoured in 1906.

Ochoa was born in Lueca, a coastal town in the Asturias region of northern Spain. When he entered the University of Madrid in 1922, Cajal had just retired but the aura of his 'heroic research' had remained. Ochoa obtained a medical degree, then the only course generally available in Europe for a career in physiology. With the enthusiasm and energy that characterized him throughout his life, he plunged into research in physiological chemistry in the laboratory of Juan Negrin, who later became the president of the short-lived Spanish Republic.

Studies of the action of guanidine on frog melanophores and the development of a micromethod for assaying creatine in muscle, were followed by two years in Otto Meyerhof's laboratory, first in Berlin and then in Heidelberg, exploring the energetics of muscle contraction. Except for a few brief stays in Madrid, he worked abroad — on glycolysis in muscle extracts at the National Institute for Medical Research in London with Harold W. Dudley in 1932; with Meyerhof again in Heidelberg in 1936 on phosphorylation in invertebrate muscle; in Plymouth with marine biologists in 1937; on demonstrating with Sir Rudolph A. Peters in Oxford in 1938 that thiamine pyrophosphate is the vitamin B₆ coenzyme of carboxylase; and finally in St Louis, Missouri, with Carl F. and Gerty Cori in 1940, on the isolation and characterization of fructose 1-phosphate and inorganic pyrophosphate. Buffeted by turbulence and wars in Spain, Germany and Britain, his steadfastness and

devotion to science were severely tested, but he never wavered.

When I joined him as a postdoctoral fellow in 1946, he was assisted by one technician and a graduate student (Alan H. Mehler) in borrowed space in the department of biochemistry in the New York University College of Medicine, having been



Severo Ochoa, 1956.

evicted from the psychiatry department. He was pursuing what I regarded to be the holy grail of biochemistry — the soluble enzymes that make ATP upon oxidation of carbon compounds in the citric acid (Krebs) cycle. Although the key event of ATP synthesis was later shown to reside in the indissoluble mitochondrion, Ochoa did more than anyone else to resolve the enzymes and clarify the mechanisms that generate and use citric acid at the hub of intermediary metabolism. Insights and techniques that emerged from these studies led him to explore and understand stages in the synthesis of fatty acids and the conversion of light to chemical energy in photosynthesis.

Most noted of his myriad contributions was the serendipitous finding by a postdoctoral fellow, Marianne Grunberg-Manago, of an enzyme, polynucleotide phosphorylase, that made an RNA-like polymer from nucleotides. Whereas the physiological function of the enzyme proved to be the breakdown of RNA, its capacity to syn-

thesize RNAs of defined composition made it a key reagent in deciphering the genetic code. These studies set the stage for discoveries of RNA synthetases encoded by RNA viruses and the mechanisms of translational processes in protein synthesis, particularly the direction of translation, termination codons and a menagerie of initiation factors.

A courtly, charming El Greco-like figure, Ochoa played host to a legion of postdoctoral fellows, students and sabbatical guests who came from every corner of the world and left to become leaders in science. His many medals, diplomas, citations and honorary degrees are displayed in a handsome museum founded by a distinguished disciple, Santiago Grisolia, and located in Valencia near a street that bears his name (as do other streets in most cities in Spain). Don Severo Ochoa was a name and face nearly as well known to the Spanish cab driver and housewife as that of King Juan Carlos. The Severo Ochoa Centre for Molecular Biology, founded with Ochoa's guidance and patronage and directed by another disci-

ple, Margarita Salas, has become one of Europe's leading centres of molecular biology.

Upon retirement from New York University in 1974, Ochoa commuted to the Roche Institute in Nutley, New Jersey, where as a Distinguished Member he continued a vigorous programme on various aspects of protein synthesis. When the attachments to his beloved New York were weakened by the departure of most of his contemporaries in science, he and Carmen, his wife for 52 years, responded to entreaties to return to Spain in 1985. Her death soon after was a devastating loss from which he never recovered. The biochemist's biochemist, revered teacher, international citizen and loyal friend to so many has now left us too.

Arthur Kornberg

Arthur Kornberg, joint winner with Ochoa of the Nobel Prize for Physiology or Medicine in 1959, is in the Department of Biochemistry, Stanford University School of Medicine, Stanford, California 94305, USA.

The chimaeras speak again

Marc G. Caron

ONE way of finding out more about protein structure and function is to create chimaeras from different members of a particular protein family. That is what Eiselé and colleagues (page 479 of this issue¹) have done. Their results are remarkable. They created functional chimaeras between two different ligand-gated channel receptors, which is quite a feat in itself, and are able to demonstrate that the two domains concerned each have a distinct role in the operation of this receptor family.

Understanding of the protein components involved in cellular signal transduction has benefited enormously from molecular cloning of the complementary DNAs and genes involved. But few if any of these proteins are known at a level of structural resolution that allows a full appreciation of the molecular interactions between a ligand and its receptor, and of how this interaction is translated into an intracellular signal. This is where the production of chimaeras comes in. Unlike more conventional ways of mapping functional domains of proteins (such as susceptibility to proteolysis or deletion mutagenesis, which often impair or destroy the function of interest), this approach usually allows the gain or transfer of a function from one protein to another. But such marriages of convenience are not at all easy to bring about — in particular, it has proved especially difficult to create useful chimaeras of transmembrane proteins such as receptors and ion channels.

Eiselé and colleagues give an excellent account of a design strategy for such studies, and demonstrate its power. They show, quite convincingly, that a chimaeric channel protein, formed from the amino-terminal domain of the $\alpha 7$ subunit of the neuronal nicotinic acetylcholine (nACh) receptor and the putative transmembrane domains of the 5-hydroxytryptamine receptor (5HT₃), generates a current that has the pharmacological properties of the $\alpha 7$ nicotinic receptor and the channel properties of the 5HT₃ ligand-gated channel.

Domains

Although they belong to the same family, the neuronal nACh and 5HT₃ receptors are not closely related (amino-acid conservation is about 20–25 per cent). Their structural characteristics are similar, in that both have a large amino-terminal domain (presumably extracellular), followed by four putative transmembrane domains and a carboxy-terminal domain (also predicted to be extracellular). Most ligand-gated channel receptors function as

heterooligomers in tissues², but in some instances ligand-binding subunits of these channels can form functional homooligomers upon heterologous expression³. From numerous biochemical and mutagenesis studies it has been inferred that the ligand/neurotransmitter binding site for these channels resides on the extracellular face of the proteins^{4,5}, and that the transmembrane domains confer channel properties^{6–9}. But whether the ligand-binding and channel properties are independent and separable had not been established.

The $\alpha 7$ -5HT₃ chimaeric receptor created by Eiselé and co-workers is activated by micromolar concentrations of acetylcholine and nicotine and is completely blocked by nACh receptor antagonists with the same potency as the expressed wild-type $\alpha 7$ nACh receptor. 5HT is devoid of activity. In a series of intricate 'manips' (as they say in France), involving the clamping of intracellular Ca²⁺ and rapid switches of the concentrations of calcium from low to high during electrophysiological recording, the authors show that the chimaeric channel displays both the allosteric effect of calcium on the binding of acetylcholine (an effect observed with the $\alpha 7$ and not with the 5HT₃ channel) and the channel blocking ability of Ca²⁺, a property of the 5HT₃ receptor. The wild-type 5HT₃ receptor is blocked by Ca²⁺ because, unlike the wild-type $\alpha 7$, it is impermeable to calcium ions.

The significance of all this is that the authors can unequivocally assign the ligand-binding function and the Ca²⁺ allosteric site of the $\alpha 7$ to the amino-terminal segment of the nACh receptor, and the channel selectivity properties to the transmembrane core of the 5HT₃ receptor. This conclusion was in fact predictable, but the really remarkable discovery here is how functionally independent and separable these two domains are. Whatever signal is induced by occupancy of the ligand-binding domain of one receptor, the information generated is capable of activating (opening) the channel pore of a distantly related member of the ligand-gated receptor family. Although the degree of amino-acid conservation between the proteins suggests that they evolved from some common ancestral gene many millions of years ago, these results point to a remarkable conservation in their functional principles through evolution.

As opposed to the apparent dissociability of ligand binding and ion selectivity, properties of the chimaeric channel protein (such as the kinetics of current onset and inactivation or desensitization) were

intermediate between the fast $\alpha 7$ and the slow 5HT₃ kinetics. From those characteristics it seems that the two domains work in concert in determining the kinetic properties of the receptors.

Strategy

The experimental strategy used by Eiselé *et al.* is also worth highlighting. Chimaeric protein constructs have been instrumental in understanding the structure and function of many components of transduction pathways^{10–13}. But it may be that the chimaeras that have been reported represent only a few of all those that have been engineered in various laboratories. Eiselé *et al.* produced five different chimaeras by systematically moving the junction between the amino-terminal domain of the $\alpha 7$ receptor and the first transmembrane span of the 5HT₃ receptor to five distinct conserved residues along the stretch of about 50 amino acids. Only two of these constructs generated measurable currents in response to the binding of acetylcholine, even though at least four of the five chimaeras displayed ligand binding. These findings suggest that many of the chimaeric proteins produced were not compatible with function. Such a systematic approach is certainly a tedious one. But until the compatibility of various protein domains can be predicted from structural information, I think it is the best available strategy for producing functional chimaeras.

As remarkable as the properties of Eiselé and colleagues' channel are, one of the next steps should surely be to determine whether the reverse chimaeric protein or similar chimaeras between other members of this large protein family will also show similar properties. Such studies would also tell us something of the generality of the concept that domains of these channel proteins can function as autonomous units. □

Marc G. Caron is in the Howard Hughes Medical Institute Research Laboratories, Duke University Medical Center, PO Box 3287, Durham, North Carolina 27710, USA.

1. Eiselé, J.-L., Bertrand, S., Galzi, J.-L., Devillers-Thiéry, A., Changeux, J.-P. & Bertrand, D. *Nature* **366**, 479–483 (1993).
2. Hocho, F. *Eur. J. Biochem.* **158**, 211–226 (1986).
3. Couturier, S. *et al. Neuron* **5**, 845–856 (1990).
4. Dennis, M. *et al. FEBS Lett.* **207**, 243–249 (1986).
5. Changeux, J.-P., Galzi, J.-L., Devillers-Thiéry, A. & Bertrand, D. *Q. Rev. Biophys.* **25**, 395–432 (1992).
6. Hume, R. I., Dingledine, R. & Heinemann, S. F. *Science* **253**, 1028–1031 (1991).
7. Vendoorn, T. A., Burnashev, N., Monyer, H., Seeburg, P. H. & Sakmann, B. *Science* **252**, 1715–1718 (1991).
8. Sommer, B. & Seeburg, P. H. *Trends pharmac. Sci.* **13**, 291–296 (1992).
9. Galzi, J. L. *et al. Nature* **359**, 500–505 (1992).
10. Evans, R. M. *Science* **240**, 889–895 (1988).
11. Koblika, B. K., Koblika, T. S., Regan, J. W., Caron, M. G. & Lefkowitz, R. J. *Science* **240**, 1310–1316 (1988).
12. Masters, S. B. *et al. Science* **241**, 448–451 (1988).
13. Ulrich, A. & Schlessinger, J. *Cell* **61**, 203–212 (1990).

RÉSUMÉ

Fits and starts

S. T. NICHOL *et al.* have taken a fresh look at molecular evolution by sequencing *P* genes from vesicular stomatitis virus, VSV (*Proc. natn. Acad. Sci. U.S.A.* **90**, 10424–10428; 1993). This is an RNA virus which infects various farm animals and humans in Central and North America, and over a period of 50 years many diverse forms of it have been isolated. The authors sequenced the *P* gene of 149 isolates, then analysed the sequences phylogenetically. The upshot is an evolutionary tree that, say the authors, points to punctuated evolution having taken place as VSV spread from Panama up to the United States — something of a provocative conclusion.

Shuttle service

THOSE fascinating creations rotaxanes consist of a molecular 'rotor' threaded by a linear 'axle'. Synthetic chemists are dreaming up variations on the theme; last year, for instance, A. Harada *et al.* reported (*Nature* **356**, 325–327; 1992) that several cyclodextrins could be threaded on a long-chain backbone like beads on a necklace. A. C. Benniston and A. Harrison (*Angew. Chem. int. Edn Engl.* **32**, 1459–1461; 1993) are content with just one bead. But their system is photoactive: light induces the formation of a pair of radical cations, as an electron is shunted from the dialkoxybenzene thread onto the positively charged cyclophane bead. The fun starts if, instead of simple recombination of the unpaired electrons, the dialkoxybenzene radical cation then transfers both its positive charge and the lone electron to one of the bulky ferrocene 'stopper' units at its ends. Electrostatic repulsion between this unit and the bead repels the bead, which shuttles to the opposite end of its thread.

Knockouts ok

SYNAPTIC vesicles are specialized organelles which rapidly exocytose in response to a depolarization-induced rise in intracellular calcium; this process may be regulated in part by the phosphorylation state of the synaptic vesicle protein, synapsin I. Now, T. Sudhof and collaborators (*Cell* **75**, 661–670; 1993) have bred synapsin I knockout mice which, surprisingly, seem to have no serious impairment of the nervous system. The mice do show an increase in paired pulse facilitation, a form of short-term plasticity in which more neurotransmitter is released after a second stimulus than after the first, indicating that in normal mice synapsin I may limit the amount of neurotransmitter released in response to repeated stimuli. Another protein may be compensating for synapsin I in the knockout mice, possibly its relative, synapsin II. It remains to be seen how well mice manage if both synapsins are knocked out.

PLANETARY SCIENCE

Top-down convection

Norman H. Sleep

MANY natural systems involve cooling of an initially hot, nearly isothermal fluid from the top down. For example, lava lakes and magma chambers are often cooled by proximity to the Earth's surface, and the oceanic lithosphere cools downward as it spreads away from the mid-ocean-ridge axis. Quantitative studies of the process have been surprisingly rare, but modelling by Worster and co-workers¹ and a stickier study by Davaille and Jaupart² are starting to fill the gap.

At first sight, the process seems simple.

But the topic has led to controversy rather than agreement. It is obvious that only temperature contrasts in the actively flowing part of the magma chamber (not the much larger contrasts in the rigid lid) drive convection. On the one hand, it may be argued that magmas are quite fluid (like thick soup) and thus should convect if any unstable temperature gradient is present. On the other hand, cooling of a magma causes crystallization so that any temperature decrease will freeze the magma, precluding convec-

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Hot lava erupts in Hawaii — but what goes on beneath the surface?

The strong temperature dependence of viscosity for materials typical of the Earth causes a cool rigid lid to develop over a hot interior that is fairly fluid. The thickening of the lid and the cooling of the hot interior with time depend strongly on whether thermally driven convection in the hot interior can transfer heat. Two extreme cases illustrate this point. At one extreme, conduction alone occurs. The lid thickens in proportion to the square root of time and an interior point remains at its initial temperature until cooling penetrates from the surface. At the other end of the scale, the interior convects vigorously so that the heat supplied by convection to the base of the lid balances the heat lost by conduction out of its top. The entire interior thereby supplies heat to the surface by cooling slowly. The lid gradually thickens as the cooling interior becomes more viscous and convects less vigorously.

The existence of convection in magma chambers and its vigour have attracted the interest of petrologists because the composition of erupted magmas as well as of crystallized igneous rocks is strongly

affected. So it is encouraging to see that progress has been made on two approaches to this problem.

Worster and co-workers¹ base numerical models on the kinetic undercooling of magmas: that is, the finite temperature difference between the equilibrium temperature for slow crystallization in the interior of the magma and the actual temperature at which crystals can grow rapidly on the roof of the chamber. Their experiments indicate that this difference is a few kelvin — not much, but enough to drive vigorous convection. Worster *et al.* also obtain a simple expression for the convective heat flow at the top of the magma chamber

$$q = 0.14 k \Delta T \left[\frac{\alpha \rho g \Delta T}{\kappa \eta} \right]^{1/3}$$

where 0.14 is an empirical constant, k is thermal conductivity, ΔT is the kinetic temperature difference, α is the thermal expansion coefficient, ρ is density, g is the acceleration due to gravity, κ is thermal diffusivity, and η is the viscosity of the

interior. This expression has the property (shared with many other results relating to convection where inertial forces are not significant) that the heat transfer does not depend on the thickness of the body, provided that the active boundary layer is thin in comparison. It allows calculation of convective heat transfer without having to solve a fluid dynamics problem. It also allows examination of features of petrological interest. For example, in addition to solidifying from the top, a magma chamber fills from the bottom with crystals formed within the chamber. The relative contribution of each filling mechanism (observed in exposed igneous bodies) can be calculated, as the amount of interior crystallization is related to cooling of the chamber interior by convection.

Davaille and Jaupart^{2,3} have been doing experiments with ordinary golden syrup, the viscosity of which changes gradually with temperature. By measuring temperature and heat transfer within the interior of the fluid, they obtained a very similar equation, with the modifications that the constant is 0.47 and that

$$\Delta T \equiv \frac{-\eta}{\delta\eta/\delta T}$$

evaluated at the interior temperature (or crudely the temperature interval to change viscosity by a factor of e). Typically, this ΔT is larger than the ΔT associated with kinetics as noted by Worster and co-workers. Davaille and Jaupart found that viscosity varies by a factor of 10 across the active part of the boundary layer.

It is not clear which formulation better represents the physics of real igneous bodies, as both models give qualitatively similar predictions. For example, both groups obtained reasonable simulations of the Makaopuhi lava lake on the island of Hawaii. Investigations of melts in the laboratory will not fully resolve the issue as grain-scale physics are likely to differ among such systems. Rather, detailed examination of the chemical composition and fabric of material from the roofs of magma chambers is warranted.

Davaille and Jaupart's formulation is directly applicable to convection at the base of planetary lithospheres. Such convection is a possible mode of heat transfer within the Earth and the only significant mode of convective heat transfer within the one-plate planets (the Moon, Mercury, Mars and maybe Venus). It has been suspected for some time that such convection carries little heat into the lithosphere of the Earth because there are no evident topographical and gravity anomalies that can be associated with it⁴. (A possible exception is near the East Pacific Rise where the mantle is believed to be hotter than normal.) This inference is confirmed by applying Davaille and Jaupart's form-

ula to obtain the result that a viscosity of 10^{17} Pa s is needed if convection is to balance heat flow through older oceanic crust (50 mW m^{-2} , assuming that $\Delta T = 40 \text{ K}$; the remaining parameters are well constrained). This viscosity is too low to supply a drag force at the base of plates, so it can probably be excluded. The Earth's mantle would need to be $100\text{--}200 \text{ K}$ hotter than at present to have this viscosity at the base of the lithosphere. Another possibility is that ΔT is around 100 K and convection is vigorous (Davaille and Jaupart, personal communication).

An additional implication is that the interior of a one-plate planet needs to be fairly fluid and thus hotter than the Earth's interior if convection beneath the lithosphere is to remove interior heat. Venus is particularly interesting, and Turcotte⁵ recently proposed that plate tectonics are episodic there. The hot surface temperature allows plate tectonics to transfer heat efficiently and cool the interior. Eventually the interior becomes too cool and too viscous, and plate tectonics halt. This last happened about 500 million years ago. Convection at the base of the lithosphere is at first too sluggish to prevent the lithosphere from cooling

downwards by conduction while the interior heats up by radioactivity. It may take several hundred million more years for the interior of Venus to become hot enough that convection can begin to thin the lithosphere again, and even longer to restart plate tectonics.

Historically, convection cooled from above has received much less laboratory and mathematical attention than convection between two isothermal surfaces. This situation has now changed; understanding has now progressed to the point where quantitative calculations can be made for problems of geological interest. There is plenty to occupy scientists interested in magma bodies and the thermal evolution of planetary interiors. □

Norman H. Sleep is in the Department of Geophysics, Stanford University, Stanford, California 94305, USA.

1. Worster, M. L., Huppert, H. E. & Sparks, R. S. J. *J. geophys. Res.* **98**, 15891–15901 (1993).
2. Davaille, A. & Jaupart, C. *J. Fluid Mech.* **253**, 141–166 (1993).
3. Davaille, A. & Jaupart, C. *Geophys. Res. Lett.* **20**, 1827–1830 (1993).
4. Davies, G. F. *J. geophys. Res.* **93**, 10481–10488 (1988).
5. Turcotte, D. L. *J. geophys. Res.* **98**, 17061–17068 (1993).

PALAEOCLIMATOLOGY

Ancient tropical methane

F. A. Street-Perrott

"One minute it darted off like a kingfisher, and the next it entirely disappeared. At times it grew as big as an ox's head, and then straightaway shrank to a cat's eye . . . finally . . . it returned to frisk in the reeds."

So George Sand, in 1848, described a will o' the wisp — a puff of ignited methane — hovering over a marsh in central France¹. Such emissions must have had a long geological history, and on page 443 of this issue, Chappellaz *et al.*² present a detailed record of atmospheric methane extracted from gas bubbles entombed in the Greenland ice sheet.

The record was obtained by the Greenland Ice-core Project (GRIP) research team, which has drilled through over 3 km of ice beneath Summit station³. It spans the period 40,000–8,000 years BP (calendar years before AD 1950). This section of the Summit ice core exhibits no fewer than 11 abrupt fluctuations in methane, which the authors attribute to variations in the rate of production by fermentation of plant biomass in tropical wetlands. The fluctuations point to the existence of strong feedback between the biosphere and climate on a timescale of centuries².

The broad trends seen in methane parallel those previously discovered at Vostok station in Antarctica⁴ and elsewhere. They include a doubling from 350

p.p.b.v. (parts per billion by volume) in glacial times to 700 p.p.b.v. during the present interglacial, and an earlier interval of high mean concentrations around 35–30 kyr BP. A great advantage of this new study, however, is its excellent time resolution, which ranges from 100–200 years in the upper part to 600–900 years at 40 kyr BP. This reflects the high accumulation rate in Greenland (0.21 m of snow per year) and the close sampling interval used (10 m spacing from 1,330 to 2,270 m depth)². Sadly, the upper 1,330 m of ice was fractured and unsuitable for measurement of trace gases, but the core quality below that depth was excellent.

The methane record from Vostok station shows clear periodicities of 110, 38, 24 and 19 kyr (ref. 4), suggesting that the long-term trends in methane were driven by variations in insolation caused by the Earth's changing orbital geometry. It is much harder to explain the rapid fluctuations seen at Summit station.

A key finding is that the methane peaks at Summit parallel the interstadials (abrupt climate warmings over Greenland) evident in the $\delta^{18}\text{O}$ curve³ for the same core. The two curves are in phase within $\pm 100\text{--}200 \text{ yr}$ (ref. 2). This observation adds weight to the claim that the underlying climate changes were hemis-

pheric or even global in extent. The direct radiative impact of the abrupt methane changes would have been a warming of a tenth of a degree at most, so they are not enough to cause the climate changes, but the fact that the two are in phase suggests that they have a common cause.

Three main hypotheses have so far been advanced to explain the variations in atmospheric methane during the past 150 kyr. These all involve variations in methane sources, as model calculations suggested that the principal sink, atmospheric hydroxyl radicals, was slightly larger during the last glacial period⁵, but that



The tall C₄ sedge *Cyperus papyrus* on Rurie Swamp, Kenya. Papyrus swamp has very high primary productivity and strongly reducing muds, which favour high rates of methane generation.

the change was too small to explain the observed decrease in methane.

The first hypothesis attributes the changes to the spread of high-latitude peatlands, which emit methane, at the end of the last glaciation and in other, earlier, periods of high summer insolation and warm temperatures at high northern latitudes⁶. But Chappellaz and colleagues' record shows that methane increased to Holocene levels while ice still covered many modern peatland areas, so this hypothesis appears unlikely. A second attributes the rapid rise in methane during deglaciation to catastrophic outgassing from methane hydrates (clathrates) stored at high pressures and low temperatures under the Pleistocene ice sheets, in permafrost, and within the marine sediments of the outer continental shelves⁷. Chappellaz *et al.* play down this explanation on the grounds that no methane spikes of the required magnitude (several 10³ p.p.b.v.) occurred during deglaciation, but a slower

and more modest contribution from this source cannot be ruled out. Third, the close correspondence between major increases in methane in the Vostok ice core and periods of greatly enlarged lakes in the Sahara has been cited as evidence for a tropical wetland source⁸.

The detailed form of the methane curve for the past 15 kyr provides strong support for the tropical hypothesis. It shows sharp rises at 14.45 and 11.55 kyr BP, with minima during the Younger Dryas cold snap (12.7–11.55 kyr BP) and around 8.2 kyr BP. Severe droughts in the northern tropics⁹ culminated around 11.8 and 8.2 kyr BP. The first was registered in northwest, west and east Africa, Tibet and northern South America, and the second in at least the first four of these areas. Both of these methane minima correlate with cold stadials in the Summit ice core². They also correspond to episodes of marked freshening of the northern North Atlantic¹⁰, suggesting that both cooling in Greenland and weakened tropical monsoons, with their concomitant drying of wetland and decrease in methane emissions, were responses to large-scale changes in the Atlantic circulation. Specifically, it is thought that the rate of production of North Atlantic Deep Water declined in the Greenland-Norwegian and Labrador seas.

If this hypothesis is correct, then another sharp decrease in methane should have occurred around 5,000 yr BP, when many tropical lakes dried up⁹, but the evidence for this would lie in the top, brittle section of the ice core. Another test, which would rule out one of the alternative theories, would be radiocarbon-dating of the fossil methane using accelerator mass spectrometry. Emissions from clathrates should be of old methane, whereas tropical methane should be more nearly contemporary with the date of bubble closure. True, this would be technically difficult — but no will o' the wisp is easy to track down. □

F. A. Street-Perrott is in the Environmental Change Unit and School of Geography, University of Oxford, Oxford OX1 3TB, UK.

1. Sand, G. *La Petite Fadette* (Garnier Flammarion, Paris, 1967).
2. Chappellaz, J. *et al.* *Nature* **366**, 443–445 (1993).
3. Dansgaard, W. *et al.* *Nature* **364**, 218–220 (1993).
4. Chappellaz, J. *et al.* *Nature* **345**, 127–131 (1990).
5. Thompson, A. M. *Science* **256**, 1157–1165 (1992).
6. Crowley, T. J. *Nature* **353**, 122–123 (1991).
7. Nisbet, E. G. *J. geophys. Res.* **97**, 12859 (1992).
8. Petit-Maire, N. *et al.* *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **86**, 197–204 (1991).
9. Street-Perrott, F. A. & Roberts, N. in *Variations in the Global Water Budget* (eds Street-Perrott, A., Beran, M. & Ratcliffe, R.) 331–345 (Reidel, Dordrecht, 1983).
10. Duplessy, J.-C. *et al.* *Nature* **357**, 485–487 (1992).

DAEDALUS

Soular mass

FROM time to time the claim is made that a dying human being loses weight, as his soul departs. Losses from 10 mg to 1.8 g have been quoted. In fact, the gas exchange and internal motions of a body, alive or dead, prevent such accurate weighing. Daedalus has a better way.

The subject must be sealed in an enclosed cell equipped with an air-regenerator, supported on sensitive piezoelectric transducers and fitted with the finest inertial-navigation accelerometers. He must be fitted with sensors for heart-beat, breathing, peristalsis and other bodily movements. A central computer will receive all the transducer and sensor signals, and will refine a model of the whole system and generate its equation of motion. It will validate its findings by applying small precisely known forces and torques to the system, via the transducers. By steadily propagating the equation forward in time, and optimizing it to fit the observed perturbations, it will soon refine its model to high accuracy. Its equation of motion will contain precise values for the mass of the system, the position of its centre of gravity, and the magnitude and direction of its principal moments of inertia.

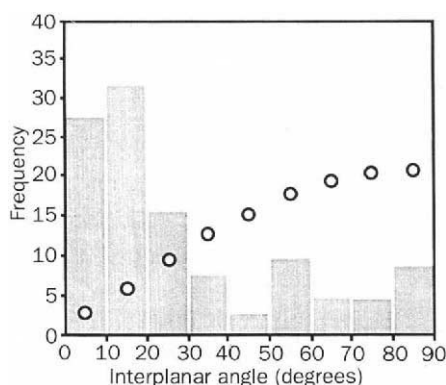
When the subject dies, this equation will cease to fit. The computer will carry on processing the transducer data, and will ultimately optimize a new equation of motion, with different mechanical constants. Propagating this backwards to the moment of death will then reveal the discontinuities caused by the loss of the soul. The change of mass will reveal its weight; the change of centre of gravity will reveal the site it had vacated; the alteration of inertial moments will reveal its mass distribution. The form of the recoil at the instant of death will reveal the velocity, direction and spin of the departing soul, and the time it had taken to leave the body.

This new science of spiritual mechanics will place a heavy burden on its experimental subjects. Only a few people, such as astronauts, have been acclimatized to close confinement in conditions of mortal danger, while wired up to a comprehensive set of invasive sensors. But even a qualified volunteer could not be asked to lie in the apparatus until he died of natural causes: still less could he be invited to commit suicide. Perhaps a condemned murderer awaiting lethal injection (as practised in American states such as Arkansas) might volunteer for the experiment, to allow science to benefit from his last ordeal. The test may have to wait till an astronaut commits murder in Arkansas.

David Jones

Amino/aromatic interactions

SIR — Rodham *et al.*¹ detailed both experimental and theoretical investigations of the structure of the ammonia/benzene complex. Following the identification of an amino/aromatic hydrogen bond in SH2 domain/peptide binding by Waksman *et al.*², there has been a resurgence of interest in such interactions in proteins. We believe that this interest needs to be accompanied by an understanding of the rarity and relatively weak energies of these interactions compared with other kinds. Before Levitt and Perutz³ first



Distribution of the interplanar angle between sp^2 nitrogen and an aromatic ring for 116 amino/aromatic interactions in which the nitrogen lies above the ring face. This is defined by requiring the nitrogen to be within 20° of being directly above either a ring carbon or the ring centre, and no more than $3.8 \text{ \AA}$ away from the same site. The dataset includes both side- and main-chain nitrogens and both phenylalanine and tyrosine rings. Seven interactions involving lysine side-chains are, however, excluded from this plot as the interplanar angle is undefined, these nitrogens being sp^3 hybridized. Our definition of an amino/aromatic hydrogen bond requires an 'above ring' interaction with $N-H \cdots C < 120^\circ$ and an interplanar angle (not applicable to lysine side-chains) $\geq 30^\circ$. All 'above ring' interactions with interplanar angles $< 30^\circ$ are classified as stacked. Circles represent the sinusoidal expected distribution.

suggested that amino/aromatic hydrogen bonds might be important in proteins, Legon and Millen⁴ had already established that proton donor/ π -electron interactions are, at least for small molecules, observed if, and only if, conventional hydrogen bonding to lone-pair acceptors is not possible. Indeed, our own energy calculations reinforce the view that

such an interaction is significantly weaker than a normal hydrogen bond, but preferable to leaving a potential hydrogen-bonding proton totally unfulfilled.

We have studied various amino/aromatic interactions, especially those 'above ring' cases where any nitrogen atom lies above the phenylalanine or tyrosine rings (requiring an atom-atom contact distance of less than $3.8 \text{ \AA}$). We used 58 non-homologous high-resolution Brookhaven⁵ protein structures and added inferred hydrogen positions to the structures, enabling us to search for all amino/aromatic hydrogen bonds. However, we found that the majority of 'above ring' amino/aromatic interactions involve stacking of the planar sp^2 nitrogens above their aromatic partners; only a few are hydrogen bonded. Out of 1,171 residue/residue interactions, 123 are 'above ring'; of these, 36 are hydrogen bonded, 76 stacked and 11 neither (see figure). Thus only 3 per cent of all amino/aromatic interactions found are hydrogen bonds.

Our *ab initio* calculations find that the isolated dimer interaction energy, when taken alone, favours hydrogen-bonded conformations over stacked ones. Within proteins, in contrast, the stacked geometry is generally favoured as it allows all potential hydrogen-bonding protons to interact with conventional acceptors,

whether from protein or solvent. Thus the forming of conventional hydrogen bonds generally dominates amino/aromatic interactions, as the relative paucity of amino/aromatic hydrogen bonds shows.

Where amino/aromatic hydrogen bonding is indicated, we usually find that one proton points roughly towards a ring carbon; the second proton on an amino group points away, apparently in accordance with the monodentate nature of the ammonia/benzene¹ and water/benzene⁶ structures. But as this second proton generally forms a conventional hydrogen bond of its own, the geometry is in fact likely to be a compromise between optimizing the two donor/acceptor interactions⁷.

John B. O. Mitchell, C. Lillian Nandi, Shamimara Ali, Ian K. McDonald, Janet M. Thornton

Biomolecular Structure and Modelling Unit, Department of Biochemistry and Molecular Biology, University College London, Gower Street, London WC1E 6BT, UK

Sarah L. Price
Department of Chemistry, University College London, London WC1H 0AJ, UK

Juswinder Singh
Parke-Davis Pharmaceutical Research Division, Warner-Lambert Company, Ann Arbor, Michigan 48105-2430, USA

X-rays from a gamma-ray repeater?

SIR — Kulkarni and Frail¹ pointed out the occurrence within the localization box of the soft γ -ray repeater SGR1806–20 (refs 2, 3) of the 'amorphous radio nebula' G10.1–0.3 from Green's supernova remnant (SNR) catalogue⁴. They argue that all SGRs are probably young neutron stars still embedded in the remnants of the supernovae that gave rise to them, and that the characteristic signatures of SNRs and rotating, magnetized neutron stars should therefore be observable. Further radio observations of G10.1–0.3 by Kulkarni *et al.*⁵ reveal a compact nebula with its centroid at coordinates (1950) 18 h 05 min 41.76 s, $-20^\circ 25' 13''$ superimposed on an extended plateau of emission. This structure suggests a pulsar-powered SNR — a plerion — which, by analogy with the Vela pulsar⁶, could be a detectable X-ray source.

I observed the field containing the localization box of SGR1806–20 (including the smaller, revised localization box cited as a personal communication in ref. 1) with the position-sensitive proportional counter of the Rosat X-ray telescope in April 1993. The observation achieved 9,729 s at the centre of the 2° field and

yielded twelve X-ray sources within a circle of radius 0.25° centred on G10.1–0.3. Only one source of the twelve, at coordinates (1950) 18 h 05 min 41.56 s, $-20^\circ 25' 3.4''$ (counting rate $1.4 \times 10^{-3} \pm 0.4 \times 10^{-3} \text{ counts s}^{-1}$ in the energy range 0.22–2.5 keV) is within the revised localization box of SGR1806–20. The X-ray source is also coincident (within the estimated positional error of 11 arcs, which includes random and systematic errors) with the compact nebula of Kulkarni *et al.*⁵. The source extraction procedure presents no evidence of any extension of the X-ray source, but note that the Vela pulsar (at 500 pc) exhibits⁶ a wide range of angular scales ($< 18''$ to $\sim 1^\circ$) of pulsar-associated X-ray features. Such features might be seen in G10.1–0.3 with a longer observation or a more sensitive X-ray telescope.

An examination of the associated optical field reveals only one potential candidate for the X-ray source. This is a reddened object close to the IIIaJ plate limit and positioned on the edge of the X-ray error box diagonally opposite the radio position reported in ref. 5. A suitably obscured low-mass X-ray binary

1. Rodham, D. A. *et al.* *Nature* **362**, 735–737 (1993).
2. Waksman, G. *et al.* *Nature* **358**, 646–653 (1992).
3. Levitt, M. & Perutz, M. F. *J. molec. Biol.* **201**, 751–754 (1988).
4. Legon, A. C. & Millen, D. J. *Acc. chem. Res.* **20**, 39–46 (1987).
5. Bernstein, F. C. *et al.* *J. molec. Biol.* **112**, 535–542 (1977).
6. Suzuki, S. *et al.* *Science* **257**, 942–945 (1992).
7. Mitchell, J. B. O. & Price, S. L. *Chem. phys. Lett.* **154**, 267–272 (1989).

(similar to Sco X-1) in the Galactic bulge region would have a visual magnitude of about 19–20 (ref. 7) and should be visible on the IIIaJ plate. However, the very low counting rate of the detected X-ray source makes it more likely that this candidate is merely a chance stellar occurrence in the error box.

The positional coincidence of this X-ray source and the newly discovered compact radio nebula in G10.1–0.3 supports the identification of the X-ray emission as from a pulsar-powered plerion which, very plausibly, is also the soft γ -ray repeater SGR1806–20.

B. A. Cooke

Department of Physics and Astronomy,
Leicester University,
Leicester LE1 7RH, UK

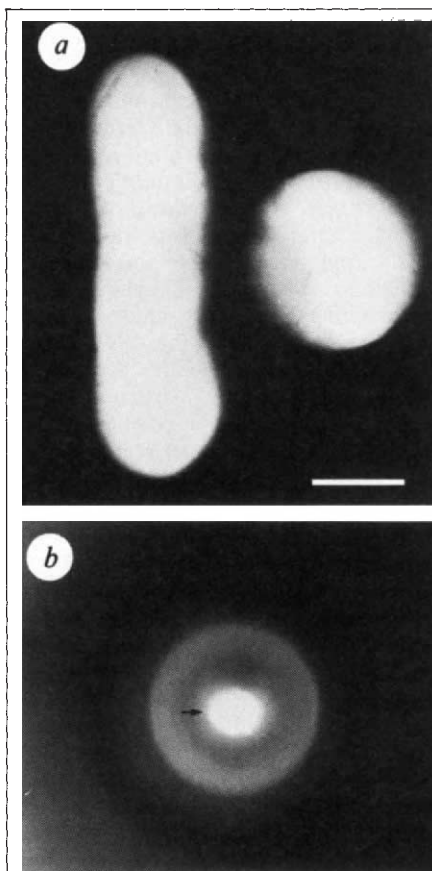
1. Kulkarni, S. R. & Frail, D. A. *Nature* **365**, 33–35 (1993).
2. Laros, J. G. et al. *Astrophys. J.* **320**, L111–L115 (1987).
3. Atteia, J.-L. et al. *Astrophys. J.* **320**, L105–L110 (1987).
4. Green, D. A. *Publ. Astr. Soc. Pacif.* **103**, 209–220 (1991).
5. Kulkarni, S. R., Frail, D. A., Kassim, N., Murakami, T. & Vasishth, G. *IAU Circ.* **5879** (1993).
6. Harnden, E. R., Grant, P. G., Seward, F. D. & Kahn, S. M. *Astrophys. J.* **299**, 828–838 (1987).
7. Cooke, B. A. & Ponman, T. J. *Astr. Astrophys.* **244**, 358–362 (1991).

Fatal attraction

SIR — We have discovered an interesting predator/prey interaction which we think is the first example of a simple bacterium harnessing the chemotaxis behaviour of another microorganism for its own ends. Cells of *Myxococcus xanthus*, a Gram-negative rod-shaped fruiting bacterium found in damp soils, on animal dung or in other habitats rich in organic matter¹, form large multicellular communities. These feed on other microorganisms, using extracellular antibiotics and degradative enzymes to digest their prey.

One of the prey microorganisms for *M. xanthus* is *Escherichia coli*, which is abundant in animal dung. *M. xanthus* moves very slowly on a solid surface by a poorly understood gliding mechanism². Wild-type *M. xanthus* cells move at a rate of 2–4 $\mu\text{m min}^{-1}$ on 1.5% agar and 5–20 $\mu\text{m min}^{-1}$ on 0.3% agar³. *E. coli* cells are propelled by flagella⁴; they move slowly on 1.5% agar and are easily overtaken by *M. xanthus*. But on softer agar (0.3%), *E. coli* swims at 900–1,200 $\mu\text{m min}^{-1}$ (ref. 4), 50–200 times faster than *M. xanthus*. While studying chemotaxis in *M. xanthus*⁵, we became concerned with how slow-moving *M. xanthus* cells can catch fast-moving *E. coli* cells. We find that *M. xanthus* cells do not show chemotaxis towards *E. coli* but instead release attractants causing *E. coli* to swim towards them in a primitive form of ensnarement.

When we plated *M. xanthus* cells (strain DZ2)³ near non-motile *E. coli* cells (strain MC4100)⁶, they did not show directed movement towards the *E. coli* colony (*a* in the figure). With longer incubation



Attraction of *E. coli* cells for its predator *M. xanthus*. *a*, *M. xanthus* (strain DZ2)³ was inoculated (10 μl containing 2×10^8 cells) in the centre of the plate, and nonmotile *E. coli* (strain MC4100)⁶ was inoculated (20 μl containing 10^8 cells) in a vertical line on the left side of the *M. xanthus* colony. The Petri plate contained CYE medium with 0.3% agar⁵. The photograph was taken after the plate was incubated at 30 °C for 24 h. *b*, *c*, *M. xanthus* (strain DZ2), pre-starved in MOPS–MgCl₂ buffer for 24 h, was inoculated (10 μl containing 2×10^8 cells) into the centre of the plates, indicated by arrows. Plates contained 30 ml 10 mM MOPS–1 mM MgCl₂, 0.3% agar and 2×10^{10} wild-type *E. coli* cells (strain AW405 for *b*) or chemotaxis mutant (strain HCB326 for *c*). Photographs were taken 2 h after the starved *M. xanthus* cells were inoculated into the *E. coli* cell suspension. Scale bars, 1 cm.

times, the colonies overlapped and the *M. xanthus* cells lysed and digested the *E. coli* cells directly in their path. Microscopic examination of the cells showed that when *M. xanthus* cells touched *E. coli* cells, the *E. coli* cells lysed in about 30 min but did not attract other *M. xanthus* cells. But when starved *M. xanthus* cells were inoculated on soft agar containing wild-type *E. coli* cells (strain AW405)⁷, the *E. coli* cells rapidly accumulated around the *M. xanthus* cells. The dense opaque zone around the *M. xanthus* colony and the clear area beyond (shown in *b* in the figure) indicate that the *E. coli* cells recognize and swim towards the *M. xanthus* colony, depleting the distant area and accumulating in the central area. We observed many *E. coli* cells in contact with the *M. xanthus* cells at the edge of the *M. xanthus* colony; the *E. coli* cells were later immobilized and digested by *M. xanthus*. A mutant *E. coli* strain deleted for the chemotaxis genes (HCB326)⁸ did not show chemotaxis towards *M. xanthus* (part *c* in the figure).

We studied this attraction behaviour quantitatively by analysing attractants in culture fluid of *M. xanthus* (see ref. 9). Attractants were first detected after *M. xanthus* had been starved for 3 hours and increased gradually with longer incubation times; the amount of attractants released was proportional to the cell density of the *M. xanthus* suspension. We repeated the capillary assays with

several characterized *E. coli* mutants with known defects in chemotaxis⁷. Strains defective in chemotaxis proteins (CheB, CheR, CheA, CheW, CheY and CheZ) no longer exhibited attraction to the *M. xanthus* culture supernatants. Strains defective in the Tsr (chemoreceptor for serine) or Tar (chemoreceptor for aspartate) showed three- to fourfold reduced response to attractants, and strains defective in the Trg (chemoreceptor for ribose and galactose) or Tap (chemoreceptor for dipeptides) showed normal attraction behaviour. Thus the chemotaxis signal-transduction machinery is required for the attraction behaviour and the attraction signals are sensed by the Tsr and Tar chemoreceptors. The attractants are probably amino acids.

Whatever the mechanism for the release of attractants, whether active or passive, the net result is that *M. xanthus* has evolved a clever method to attract *E. coli* and probably other prey microorganisms to come to them during times of starvation. The large-scale lysis of

1. Dworkin, M. & Kaiser, D. *Myxobacteria II* (American Society for Microbiology, Washington, DC, 1993).
2. Burchard, R. P. A. *Rev. Microbiol.* **35**, 497–529 (1981).
3. Shi, W. & Zusman, D. R. *Proc. natn. Acad. Sci. U.S.A.* **90**, 3378–3382 (1993).
4. Jones, C. J. & Aizawa, S. *Adv. Microb. Physiol.* **32**, 109–172 (1991).
5. Shi, W. et al. *Mol. Microbiol.* **9**, 601–611 (1993).
6. Casadaban, M. J. *J. molec. Biol.* **104**, 541–555 (1976).
7. Shi, W. et al. *J. Bact.* **175**, 5785–5790 (1993).
8. Wolf, A. J. et al. *J. Bact.* **169**, 1878–1885 (1987).
9. Adler, J. J. *gen. Microbiol.* **74**, 77–91 (1973).

the attracted *E. coli* cells and the proteolysis of their cellular contents would then produce additional attractants for more *E. coli* cells to move towards the *M. xanthus* colony. This primitive form of ensnarement must be of significant survival value for this very slow-moving microorganism.

Wenyuan Shi

David R. Zusman*

Department of Molecular
and Cell Biology,
University of California,
Berkeley,
California 94720, USA

* Author for correspondence

Dating hominid remains

SIR — McDermott *et al.*¹ are to be congratulated for obtaining mass spectrometer uranium-series dates that corroborate previously published early-uptake electron spin resonance (ESR) readings from hominid-bearing Middle Palaeolithic sites in Israel. But the authors do not mention the difficulties arising from dating isolated animal teeth derived from Garrod's excavations in the 1930s in Tabun and Skhul.

Exact recording of provenance was not practised then and understanding of site-formation processes was minimal. Unfortunately, the Skhul cave was emptied, so further dating of stratified samples cannot now be undertaken². As at Qafzeh³, all hominid burials from Skhul were uncovered from a 1-m-thick deposit (layer B2; plate I of ref. 2) covered by an additional 1 m of Middle Palaeolithic sediment (layer B1). Although McDermott *et al.* suggest that burials 1, 4 and 5 might have been later in age than the rest of the hominids, we cannot directly relate dates derived from animal teeth to those of human skeletons.

One way to resolve this problem would be direct dating of the hominid skeletal remains. Thus the younger dates $41,400 \pm 400$ and $45,500 \pm 700$ years ago may indicate the presence of a Late Middle Palaeolithic assemblage in the upper part of layer B or reflect intrusions from layer A (Upper Palaeolithic). Note that these dates fall within the range of the Early Upper Palaeolithic in the Levant⁴.

At Tabun, Garrod was also not certain about the attribution to layer C of the woman's burial (marked as Neanderthal remains in Fig. 1 of ref. 1). Even she suspected that it could have been interred by the producers of the Tabun B industry (page 64 of ref. 2). Interestingly, both in Amud and Kebara caves similarly robust skeletons were uncovered with the same industry as Tabun B (ref. 5). The main portion of the lithic assemblages from

Skhul B and the hominid bearing layers in Qafzeh are similar to each other and to Tabun C, where the isolated jaw (TII) resembles the Skhul and Qafzeh fossils referred to as early *Homo sapiens*. Thus dates in the range of 47,000/45,000 to 75,000/80,000 years ago are appropriate to the so-called 'Tabun B-type' industry and to the Levantine Neanderthals. By contrast, the early modern *H. sapiens* in Skhul, Qafzeh and Tabun C, all with similar 'Tabun C-type' artefacts, are dated to 80,000–120,000 years ago. Thermoluminescence dates on burnt flints from Skhul support the electron spin resonance readings⁵ and indicate an average age of $119,000 \pm 18,000$ years.

We agree with the conclusions¹ that early modern *H. sapiens* in the Levant were coeval with European Neanderthals, but believe that the former preceded the arrival of the latter in the near East, which can perhaps be attributed to the development of glacial conditions (Emiliani stages 5c or 4)⁷.

Ofer Bar-Yosef

David Pilbeam

Department of Anthropology,
Peabody Museum, Harvard University,
Cambridge,
Massachusetts 02138, USA

1. McDermott, F., Grün, R., Stringer, C. B. & Hawkesworth, C. J. *Nature* **363**, 252–255 (1993).
2. Garrod, D. A. E. & Bate, D. M. *The Stone Age of Mount Carmel* Vol. I (Clarendon, Oxford, 1937).
3. Vandermeersch, B. *Les Hommes Fossiles de Qafzeh, Israel* (CNRS, Paris, 1981).
4. Bar-Yosef, O. *Phil. Trans. R. Soc. B* **337**, 193–200 (1992).
5. Bar-Yosef, O. & Meignen, L. in *The Middle Paleolithic: Adaptation, Behavior and Variability* (eds Dibble, H. & Mellars, P.) 163–182 (Univ. Pennsylvania Press, Philadelphia, 1992).
6. Mercier, N. *et al. J. Archaeol. Sci.* **20**, 169–174 (1993).
7. Bar-Yosef, O. in *L'Homme de Néandertal* Vol. 3 (ed. Otte, M.) 31–38 (Université de Liège, 1988).

MCDERMOTT *ET AL.* REPLY — All of us, including Bar-Yosef¹, who work in material from old archaeological excavations, must make the best of the samples at our disposal. The long-standing debate about whether the Tabun Neanderthal should be ascribed to layer B, C or even D is impossible to resolve at this stage. We² have preferred to stick with the stratigraphical assignment of Garrod³. Even if the skeleton came from a different horizon (layer B or D in modern stratigraphies), so would the bulk of the associated faunal remains in Garrod's collection, and for this reason our uranium-thorium age results for the skeleton are robust.

As for Skhul, the three youngest uranium-series ages ($40,000\text{--}45,000$ years) almost certainly underestimate the true age because of a complex uranium-uptake history (they are younger than even the early-uptake ESR ages), and as we stated in our paper², further research is required to investigate whether or not two age populations are present.

We agree with Bar-Yosef and Pilbeam

that direct dating might be a solution, but it is unclear how they propose to do this. We are now tackling the problem indirectly by uranium-series and ESR dating of elements of the fauna actually associated with Skhul 5 and 9, and in the future we shall attempt direct γ -ray dating⁴ on the hominid remains.

F. McDermott*

C. J. Hawkesworth

Department of Earth Sciences,
Open University,
Milton Keynes MK7 6AA, UK

R. Grün

Quaternary Dating Research Centre,
Australia National University,
GPO Box 4, Canberra 2601, Australia

C. B. Stringer

Department of Palaeontology,
Natural History Museum,
Cromwell Road,
London SW7 5BD, UK

*Present address: Department of Geology, University College, Dublin, Belfield, Dublin 4, Ireland.

1. Bar-Yosef, O. in *L'Homme de Néandertal* vol. 3 (ed. Otte, M.) 31–38 (Université de Liège, 1988).
2. McDermott, F., Grün, R., Stringer, C. B. & Hawkesworth, C. J. *Nature* **363**, 252–255 (1993).
3. Garrod, D. A. E. & Bate, D. M. *The Stone Age of Mount Carmel* vol. 1 (Clarendon, Oxford, 1937).
4. Yokoyama, Y., Falguères, C. & Bibron, R. in *L'Homme de Néandertal* vol. 1 (ed. Otte, M.) 135–141 (Université de Liège, 1988).

Chemocline of the Black Sea

SIR — Sinninghe Damsté *et al.*¹ have investigated the question of anthropogenic impacts on chemocline depth fluctuations in the Black Sea by examining the molecular record of photosynthetic bacterial pigment degradation products in Black Sea sediments. They interpret the historical presence of anoxygenic photosynthesis to mean that "the penetration of the photic zone by anaerobic waters is not a recent phenomenon", and suggest that "natural causes for shoaling of the chemocline are more likely than anthropogenic ones". The issue of an anthropogenic influence on the chemocline affects the coastal zones of countries surrounding the Black Sea. But in my view Sinninghe Damsté *et al.* misstate several earlier arguments and, more important, rely on a small and irrelevant dataset.

Murray *et al.*² reported the rapid, unexpected rise of the chemocline by comparing hydrographic data collected in 1969 and 1988. They concluded that "Natural, interannual or decadal variations in climate and river runoff" drive vertical fluctuations in the chemocline, and argued that any effects caused by anthropogenic influences on freshwater inputs could be "superimposed on these natural variations". Water mass balance

calculations and physical forcing models show that anthropogenic perturbations are of sufficient magnitude to influence chemocline fluctuations. Contrary to the claim of Sinninghe Damsté *et al.*, it is not generally assumed that the rise in the chemocline occurred only recently and is without historical precedent. It is widely recognized that the mixed-layer depth and hydrographic properties of the Black Sea are strongly influenced by interannual climate variations. Physical oceanographic models and sedimentological evidence suggest that the chemocline has undergone large vertical fluctuations in the recent past. For example, lithogenic accumulation rates vary by nearly an order of magnitude throughout Unit II sediments and are correlated with varve thickness in Lake Saki (Crimea), indicating that large, climatically forced changes in freshwater input and water mass balance were a feature of Black Sea hydrography throughout the Holocene³. The issue raised by Murray *et al.* was whether anthropogenic perturbations have accentuated the frequency and amplitude of these natural variations.

What is required are high-resolution sediment analyses⁴. If annual changes in pigment deposition could be measured, then the frequency and perhaps the amplitude of the most recent rise in the chemocline could be compared to historical chemocline depth fluctuations^{4,5}. The sampling interval for this type of study needs to be as small as possible, and certainly less than 10 years. The few, widely spaced samples, each covering 100–1,000 years of deposition, in ref. 1 are not relevant to the issue of anthropogenic influences on the chemocline, which could only have occurred in this century. The historical presence of anoxic photo-synthesis and a shallow chemocline in the Black Sea over longer timescales does not make an anthropogenic influence on the most recent shoaling of the chemocline any less likely. High-resolution analyses of sediments at a 5–7-year sampling interval covering the past 1,500 years have yielded only ambiguous results, and it is unlikely that anthropogenic effects on the chemocline can be monitored by such sedimentological analyses⁴. Intensive time-series hydrographic studies such as those currently being made as part of the HydroBlack programme are needed to clarify the issue.

Daniel Repeta

Woods Hole Oceanographic Institution,
Woods Hole, Massachusetts 02543, USA

SINNINGHE DAMSTÉ *ET AL.* REPLY — Our paper¹ is about sulphur-bound biomarkers and their isotope compositions, a line of enquiry that we have been following for some time (see refs 6, 7). The Black Sea, with its high sulphide content and photosynthetic bacteria, was a logical site to continue this work. As we hoped, very significant quantities of isotopically and structurally distinctive materials were recovered when desulphurization techniques were applied. These techniques indicated that green photosynthetic bacteria had flourished in the past and, therefore, that anthropogenic effects are not a prerequisite for a shallow chemocline. This is what we said.

In fact, the strongest statement we can find about the timing of the shoaling of the chemocline is not in our paper but in Repeta's⁴: "Anoxic photosynthesis is most likely a recent phenomenon in the Black Sea initiated by a shallowing of the chemocline over the past decade and the subsequent development of a suboxic layer". Our results show this 'suggestion' to be incorrect. We offered these data not as an attack on Repeta's work, but as a refinement of available information.

Why has the chemocline risen? We show only that anthropogenic factors are not decisive, and this does not prove that they have no role in the present episode. Our samples were chosen to allow examination of sulphur-bound biomarkers, not to investigate short-term environmental changes. We fully agree that high-resolution studies are required for the latter purpose.

Jaap S. Sinninghe Damsté

Jan W. de Leeuw

Netherlands Institute of Sea Research
(NIOZ),

Division of Marine Biogeochemistry,

PO Box 59,

1790 AB Den Burg, Texel,

The Netherlands

Stuart G. Wakeham

Skidaway Institute of Oceanography,

PO Box 13687,

Savannah,

Georgia 31416, USA

J. M. Hayes

Biogeochemical Laboratories,

Departments of Geological Sciences and of
Chemistry,

Indiana University, Bloomington,

Indiana 47405, USA

Math E. L. Kohnen

Koninklijke/Shell Exploratie en Productie

Laboratorium,

PO Box 60,

2280 AB Rijswijk, The Netherlands

1. Sinninghe Damsté, J. S., Wakeham, S. G., Kohnen, M. E. L., Hayes, J. M. & de Leeuw, J. W. *Nature* **362**, 827–829 (1993).

2. Murray, J. W. *et al.* *Nature* **338**, 411–413 (1989).

3. Hay, B. *Paleoceanography* **3**, 491–508 (1988).

4. Repeta, D. J. & Simpson, D. J. *Deep Sea Res.* **38**, S969–S984 (1991).

5. Repeta, D. J., Simpson, D. J., Jorgenson, B. B. & Jannasch, H. W. *Nature* **342**, 69–72 (1989).

1. Sinninghe Damsté, J. S., Wakeham, S. G., Kohnen, M. E. L., Hayes, J. M. & de Leeuw, J. W. *Nature* **362**, 827–829 (1993).

2. Kohnen, M. E. L., Sinninghe Damsté, J. S. & de Leeuw, J. W. *Nature* **349**, 775–778 (1991).

3. Kohnen, M. E. L. *et al.* *Science* **256**, 358–362 (1992).

4. Repeta, D. J. & Simpson, D. J. *Deep Sea Res.* **38**, S969–S984 (1991).

Quantitative PCR

SIR — In their Product Review¹ on the competitive polymerase chain reaction (PCR), a technique that compares the accumulation of two products derived from a known amount of standard and the target, respectively, to quantitate the initial amount of target, Siebert and Larrick claim that "the initial ratio of target-to-competitor cDNA remains constant throughout the amplification", and that it is therefore "not necessary to obtain data during the exponential phase of the reaction".

On inspecting the available literature, it turns out that this point has rarely been investigated thoroughly². Some authors who have tested this claim experimentally find that the two products accumulate with different amplification efficiencies^{3,4}, so that quantification of target would lead to erroneous results if these different kinetics were not taken into account. In many systems, products accumulate up to the same concentration, irrespective of the initial amount of target. This plateau concentration, which is approximately 10^{-7} M⁵, is of course reached at later cycles if the initial amount of target is lower^{6–10}. The plateau concentration is reached when the concentration of *Taq* polymerase becomes rate-limiting, when reaction by-products have accumulated or, most probably, when PCR products successfully compete with primers during the annealing step.

The reason for the discrepancy in the kinetics of product accumulation in different competitive PCR systems^{2,4} is unclear. Therefore, when performing quantitative PCR protocols, we recommend that quantitative analysis is restricted to the exponential phase of product accumulation, as has been suggested^{11,12}, or at least that the time course of accumulation of both products is carefully established when setting up a new system.

Rudolf J. Wiesner

Bea Beinbrech

J. Caspar Rüegg

Department of Physiology II,

University of Heidelberg,

69120 Heidelberg, Germany

1. Siebert, P. D. & Larrick, J. W. *Nature* **359**, 557 (1992).

2. Bouabouia, M. *et al.* *J. Biol. Chem.* **267**, 21830 (1992).

3. Volkenandt, M. *et al.* *Proc. Soc. exp. Biol. Med.* **200**, 1–6 (1992).

4. Horikoshi, T. *et al.* *Cancer Res.* **52**, 108–116 (1992).

5. Bloch, W. *Biochemistry* **30**, 2735–2747 (1991).

6. Hoof, T., Riordan, J. R. & Tümmeler, B. *Analyt. Biochem.* **196**, 161–169 (1991).

7. Wiesner, R. J., Rüegg, J. C. & Morano, I. *Biochem. Biophys. Res. Commun.* **183**, 553–559 (1992).

8. Wiesner, R. J. *Nucleic Acids Res.* **20**, 5863–5864 (1992).

9. Nakayama, H., Yokoi, H. & Fujita, J. *Nucleic Acids Res.* **20**, 4939 (1992).

10. Peuker, H. & Pette, D. *FEBS Lett.* **318**, 253–258 (1993).

11. Wang, A. M., Doyle, M. V. & Mark, D. F. *Proc. natn. Acad. Sci. U.S.A.* **86**, 9717–9721 (1989).

12. Murphy, E. D. *et al.* *Biochemistry* **29**, 10351 (1990).

Master of metabolic cycles

John T. Edsall

Hans Krebs: Volume 1: The Formation of a Scientific Life, 1900–1933. By F. L. Holmes. Oxford University Press: 1991. Pp. 491. **Volume 2: Architect of Intermediary Metabolism, 1933–1937.** By F. L. Holmes. Oxford University Press: 1993. Pp. 481. £40.00, \$49.50 each.

HANS Adolf Krebs (1900–81) was one of the world's great investigators of intermediary metabolism. He discovered two major cyclic processes in living organisms: first the process in the liver that produces urea, and later the crucially important citric acid (or Krebs) cycle, which carries the oxidation of foodstuffs to its final step — the release of carbon dioxide. He was the first to demonstrate the existence of such cycles. In both these great discoveries, however, he reached the concept of a cyclic process only after a difficult struggle to understand experimental data that had initially baffled him.

Frederic Lawrence ("Larry") Holmes is known for many contributions, but especially for two previous books: *Claude Bernard and Animal Chemistry* (Harvard University Press: 1974) and *Lavoisier and the Chemistry of Life* (University of Wisconsin Press: 1985). In his great studies of these two scientists he made intensive use of evidence from their laboratory notebooks, in conjunction with all other available evidence, to trace the complex course of research that led to major discoveries.

In these two volumes on Hans Krebs, Holmes pursues the same aims but in a much larger context. This is not only a study of Krebs's research, it is also a comprehensive biography of Krebs's personal as well as scientific life, for his first 37 years. Holmes visited Krebs repeatedly during the last five years of Krebs's life. They had long discussions on many matters, partly in connection with Holmes's close study of Krebs's laboratory notebooks.

Volume 1 covers the first 33 years of Krebs's life. It ends with his expulsion by the Nazi government, as a "non-Aryan", from his position at the University of Freiburg in 1933, and his departure to begin a new life in England. Volume 2 tells of his first four eventful years in England; a country to which he became deeply attached. It concludes in 1937 with his greatest achievement, the citric acid cycle, and his marriage to Margaret Fieldhouse, which turned out to be a very happy one. Krebs still had 44 years of active and important life ahead of him, but Holmes has chosen this as a suitable stopping point for this comprehensive study.

In the first volume he starts with an excellent 35-page survey of the state of knowledge of intermediary metabolism in

1930, when Krebs entered the field. He then introduces us to Krebs, in his boyhood and youth, and to his family. His father, Georg Krebs, was a busy physician with a large practice, in the small, beautiful city of Hildesheim; he was a strong, responsible, cultivated man, with a commanding personality. Though brought up in the Jewish faith, he became a strong German patriot, and favoured the full assimilation of the Jews into the German

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Krebs eventually won the Nobel prize in 1953 for his work on metabolic cycles.

community. In 1921 he, with his two young sons, signed an official document renouncing the Jewish faith; they declared themselves *freireligiös*. Georg's wife Alma had died in 1919. Hans was the second of three children; his sister Elisabeth (Lise) was five years older, and his brother Wolfgang (Wolf) two years younger. Wolf later became a distinguished and innovative electrical engineer. Georg Krebs, Lise and Wolf appear repeatedly throughout the story.

Hans did well in everything at school, but not supremely well. He had, at this stage, many interests, but no special interest in science. At the age of 15, however, he decided to become a doctor, like his father, whom he greatly admired and respected, although there were clearly tensions between them. He began his medical studies in Freiburg in 1919, later gaining clinical experience in Munich and Berlin. In Freiburg he undertook successful research in tissue staining, under the

supervision of the anatomist Wilhelm von Moellendorff. The project was successful and it was published in 1923 with Krebs as the sole author. He published other research papers over the next several years, mostly related to clinical research. In these years he developed important friendships with, among others, Hermann Blaschko, David Nachmansohn and Bruno Mendel.

The four years (1926–30) that he spent with Otto Warburg in Berlin were crucial for Krebs's future. Warburg was probably the greatest biochemist of the first half of the twentieth century, known for his work on biological oxidations, photosynthesis and cancer metabolism. He ran his laboratory as an autocrat: co-workers had to arrive promptly at 8 a.m. and work until 6 p.m., six days a week, as he did. He advised young researchers to focus on great central problems, as he did; to find simple and powerful techniques for studying them, and to plan and execute frequent experiments, without worrying too much about their significance. Warburg's outlook profoundly influenced Krebs, who later ran his own laboratory along much the same lines, though in a less autocratic fashion. Krebs also realized that Warburg's manometric techniques were admirably suited for his own plans to study intermediary metabolism in tissue slices; an approach in which Warburg was not interested.

In late 1929, Krebs received a shock; after more than three years, Warburg told him abruptly that it was time to move on, he must leave by April 1930. Fifty years later, talking to Holmes, Krebs said: "Warburg handled me rather fiercely in dismissing me"; and he described it as a humbling experience, which shook him deeply. He feared that Warburg doubted his ability to achieve high distinction in science.

Krebs's truly independent career began in 1931 when he moved to Freiburg, where he was free to choose his own research problems, although he still had demanding clinical duties. He undertook a major problem that had baffled many earlier investigators, the biosynthesis of urea in the liver. (Holmes also tells the earlier history.) He found a collaborator, a medical student, Kurt Henseleit, who soon became a very capable experimenter and could carry on the experiments, with instructions from Krebs, when Krebs was busy at the clinic.

The experiments on urea formation in slices of rat liver made little progress for some months. Testing a whole series of amino acids, Krebs found that just ornithine stood out; every molecule of added ornithine caused as many as 25 additional molecules of urea to be formed. It was already well known that the liver enzyme arginase acts on the amino acid arginine to yield ornithine and urea. It

Science Photo Library

took Krebs several weeks, however, to realize that the action of arginase, and the role of ornithine in greatly stimulating urea production, provided the key to the whole problem. There must be continual resynthesis of arginine from ornithine, with arginase then acting once more on arginine to produce ornithine and urea. Every time the cycle is completed another urea molecule is released. Krebs's first short report on the urea cycle, key portions of which Holmes reproduces in English for the first time, is a masterpiece of terse and compelling reasoning. It appeared in the spring of 1932.

Biochemists recognized immediately that Krebs had made a major advance. Young people sought to work with him and in normal times he might well have had a career in Freiburg, and assembled a strong research group. But Adolf Hitler became Chancellor of Germany in January 1933, and in April Krebs was evicted, as a "non-Aryan", from his post in Freiburg. The great biochemist, F. G. Hopkins, at the University of Cambridge, made plain his interest in Krebs, and though Hopkins could not yet make a definite offer, Krebs left for England in June 1933.

Volume 2 begins with Krebs's arrival in London. Hopkins had indeed found a place for him, for one year, with a modest grant from the Rockefeller Foundation, and Krebs became a welcome member of the Cambridge biochemistry department. He developed a deep affection for England and its people. Although he actively continued his research almost from the beginning, he did not publish again until 1935, when he brought out several classic papers on the metabolism of amino acids.

Hopkins was having difficulty in getting funds for Krebs in Cambridge, and Edward Wayne, professor of pharmacology at the University of Sheffield, approached Krebs with the offer of a post as lecturer there. He would have a better salary and substantially more laboratory space than in Cambridge. Although the University of Sheffield was small, with no other biochemist there, and although Sheffield was a grimy industrial city, Krebs accepted. The proximity of the beautiful Peak District, where he could go hiking, was a factor in its favour. Wayne was a steadfast and devoted supporter, and generously turned over some of his research funds to Krebs. Even when a good position suddenly became available in Cambridge, and Hopkins offered it to Krebs, he decided to stay in Sheffield.

Throughout these four years in England, Krebs received offers from Palestine of positions there, at considerably higher status and salary than he had in England. The most appealing offers came from Chaim Weizmann in Rehovot; but eventually Krebs declined them all. Clearly he was doubtful whether the actual working

conditions for research would be as favourable as those he had in England. Also his attachment to England and his friends there was a powerful and, in the end, decisive factor. His sister Lise and her husband, still in Germany, knew by 1936 that they were in danger, and finally moved to Palestine. Hans worried over his brother Wolf, and urged him to come to England, which he and his wife eventually did, also in 1936. Wolf's great talents as an engineer assured him a good position.

The scientific developments that pervade Holmes's account reached a climax with Krebs's discovery of the citric acid cycle in 1937. Many of the component reactions of the cycle had been already found by 1930; citric acid itself was known to be important in metabolism, but nobody knew just how. The discovery of the cycle was a complex drama involving a number of players besides Krebs, notably Albert Szent-Györgyi. Carl Martius in Freiburg presented vital evidence on the synthesis of citric acid, and might have discovered the cycle; but it was Krebs, with his assistant W. A. Johnson, who did the crucial experiments, showing just where, and how, citric acid fitted into the cycle. Indeed, until that final step, nobody — not even Krebs himself — had seen that a great metabolic cycle was involved. In the final chapter, "Reflections", Holmes gives a penetrating analysis of Krebs's mode of research, its strength and its limitations.

The great achievements of the early twentieth century in intermediary metabolism, culminating in the citric acid cycle, were achieved without the use of isotopic labelling, which transformed the field. At the same time that the decisive paper of Krebs and Johnson appeared, Rudolf Schoenheimer at Columbia University in New York, whom Krebs had known as a colleague in Freiburg, was introducing the use of stable isotopes in tracing metabolic sequences, with essential aid from Harold Urey. This was the beginning of a new era.

These two volumes represent an extraordinary achievement, as a biography of a man, living in a time of turmoil and upheaval, who emerged as a great scientist. The story of both the man and the science is full and rewarding. If the reader finds some passages difficult to follow, I suggest that it may be better to pass on to the next episode in the narrative, to maintain one's momentum, and return to such difficult passages later. In any case these volumes, taken together, form one of the greatest of scientific biographies. Among studies of biochemists, I know of nothing in the least comparable with Holmes's achievement here, in its depth and breadth. □

John T. Edsall is at 7 Divinity Avenue, Harvard University, Cambridge, Massachusetts 02138-2092, USA.

Time and time again

Felicity Mellor

Black Holes and Baby Universes and Other Essays. By Stephen Hawking. Bantam Press: 1993. Pp. 175. £16.99.

AFTER the book, the guide to the book, the film of the book and the biography, a second volume by Hawking himself was inevitable. The success of *A Brief History of Time* has been so remarkable that it is difficult to discuss Hawking's writing without reference to the phenomenon that he has become. To some extent the word Hawking no longer refers to the Cambridge scientist, but to some mythical figure; he has attained a type of dual existence more common among cinema idols than scientists.

In Britain, in particular, the cosmologist as cult is extraordinary, given the often-cited public disinterest in science. Various explanations have been offered: the book is bought just for show and goes unread by most; Hawking's eminence as a scientist lifts him above most other popular science writers; his fight against motor neuron disease offers readers a human interest angle. In all probability there is a little truth in each of these explanations. There is also the more disturbing possibility that the public need scientists to match some pre-existing stereotype. An image of the hero-scientist as other; distinguishable from everyone else either through eccentricity or through physical appearance. Hawking is thus re-invented as an example of disembodied brain power. This is a delicate aspect of Hawking's success to discuss while remaining politically correct. It does not help that Hawking himself shows little sensitivity to the suffering endured by others. I hope I am not alone in finding it difficult to read an author who states that "Scientists and prostitutes get paid for doing what they enjoy".

Black Holes and Baby Universes is a collection of essays and talks written over the past 16 years. There is a good deal of overlap between the essays, and those about scientific ideas cover much of the same ground as *A Brief History*. Apart from the discussions of his no-boundary proposal, this is also material covered by many other popular science books. Unfortunately Hawking does not write with the same degree of clarity as some other authors. He frequently lapses into unexplained jargon and the essays are too short to describe the ideas in the detail they require.

About a third of the book is devoted to autobiographical sketches. Again there is little that is new. He briefly discusses his

childhood, the diagnosis of his illness and how he came to terms with it, and his attitudes towards science. I was surprised at the reasons he gives for having become a cosmologist. He says he had a need "to know how things worked and to control them. . . . If you understand how the universe operates, you control it in a way." This emphasis on control, rather than on knowledge as an end in itself, is an attitude that disturbs many non-scientists. In a society increasingly aware of environmental issues, control of nature through science is often blamed for the destruction of the environment. To seek control of the

whole Universe seems like the ultimate power game.

Without the snappy title or narrative thread of *A Brief History*, it is unlikely that this book will achieve the popularity of its predecessor. But that is not to say that sales will not be high. Even Hawking's highly technical graduate text on general relativity can now be found on the popular science shelves of some bookshops. □

Felicity Mellor is in the Department of Physics, University of Sussex, Brighton BN1 9QH. UK.

Tapping into water tables

Norman Myers

Water in Crisis: A Guide to the World's Fresh Water Resources. Edited by Peter H. Gleick. *Oxford Science Publications*: 1993. Pp. 473. £40, \$55 (hbk); £19.50, \$29.95 (pbk).

It has been said that food was the crisis of the 1970s, energy of the 1980s, and that water will prove to be the crisis of the 1990s. All that is wrong with this assertion is that food and energy continue to be crises. As for water, humans doubled their consumption between 1940 and 1980, and are likely to double it again by the year 2000. Nothing to worry about of course if supplies are plentiful. But already two billion people, or almost every third person worldwide, suffer chronic water shortages.

There are severe water deficits in the Middle East, where oil is far from the most prized liquid resource and where a key factor in the peace negotiations lies with water supplies. (Israel, Jordan and the

West Bank plan to draw water from the River Jordan in amounts far beyond the river's flows.) There is acute competition for water in the fertile farmlands of the north China Plain, a region producing one-quarter of China's grain. In the United States, one-fifth of irrigated lands depend on excessive pumping of ground water, and the Ogallala Aquifer, the largest in the world, is being depleted at a rate scores of times that of natural replenishment. Within little more than 20 years, two-thirds of Africans are expected to experience 'water anaemia'.

The consequence of an ever more thirsty world is not only lack of water for agriculture, which accounts for two-thirds of our water use. It also translates into ill-health. A full 90 per cent of disease in the developing world is related to dirty water: most of the 40,000 children who die each day could be saved through better water supplies — with all that implies for motivation for family planning.

All the more welcome, then, is this

book, edited by Peter Gleick, who is an expert in the field. It contains nine review papers on water in relation to health, agriculture, hydropower, fisheries and economic development generally. All are by leading specialists such as Malin Falkenmark, Stephen McCaffrey, Sandra Postel and Igor Shiklomanov as well as Gleick himself. But the bulk of the book is made up of 217 tables of data on water resources at global, regional and national levels; consumption patterns and trends; water-related diseases and sanitation; pollution of numerous sorts; irrigated agriculture; and water law, policies and politics.

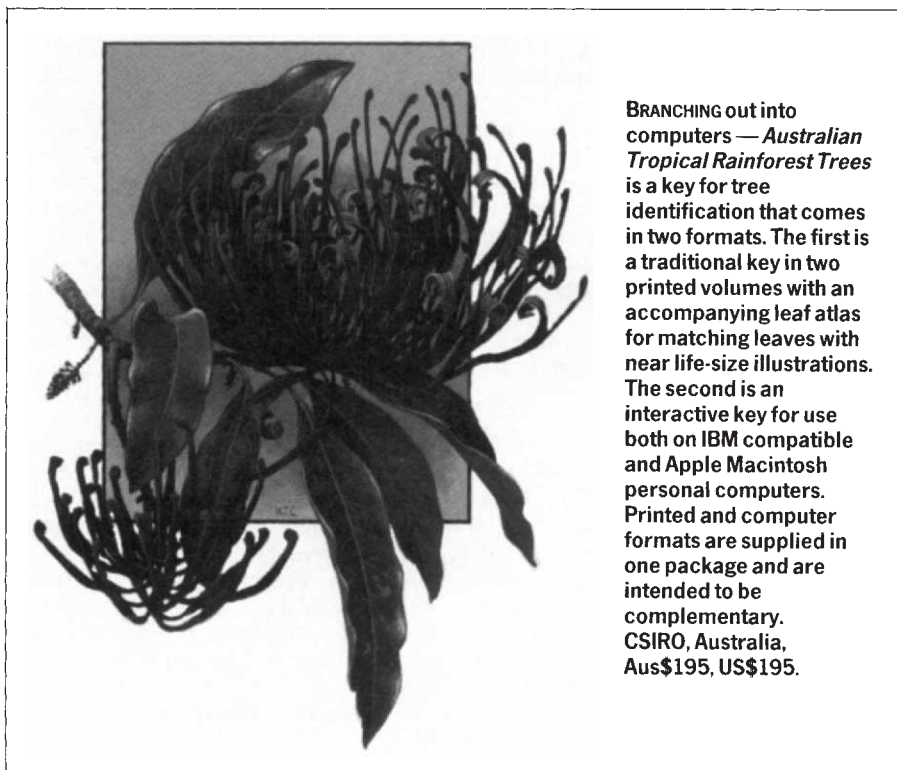
These data constitute an information stock of exceptional value, by far the best I have ever encountered. How I wish such a compendious assembly of facts and figures had been available two decades ago — in which case political leaders and policy makers might have been better aware of the emergent problem and we might conceivably have had less of a water crisis today.

The book's editor is director of the Pacific Institute for Studies in Development, Environment, and Security, based in Oakland, California. The last item, security, is especially pertinent. Gleick demonstrates that friction over water supplies offers much potential for conflict. Of 214 major river basins, three-quarters are shared between two nations and a quarter between three to ten nations. Almost half of the Earth's land surface is located within international river basins, supporting 40 per cent of the world's population; and nearly 50 nations have more than three-quarters of their territory within such areas. Tensions and violence have already erupted in the river basin of the Mekong, which is shared between Laos, Thailand, Kampuchea and Vietnam; in that of the Amur, shared between China and the former Soviet Union; and in several other river systems of geostrategic value. Of the Indus River basin, vital to the world's largest irrigation network in Pakistan, two-fifths lie in India, a nation with which Pakistan maintains relations of armed confrontation.

The book is particularly timely in that the water crisis receives all too little attention. It was hardly mentioned in the Brundtland Report or at the Rio Earth Summit. We tend to take water for granted, even though it lies at the heart not only of all life but of the livelihoods of most people on Earth. When it rains in Britain, people pull a face, whereas in Africa they dance in the streets. Two billion people will use less water for all their household needs today than will readers of this journal each time they flush the lavatory. □

David Orr

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS



BRANCHING out into computers — *Australian Tropical Rainforest Trees* is a key for tree identification that comes in two formats. The first is a traditional key in two printed volumes with an accompanying leaf atlas for matching leaves with near life-size illustrations. The second is an interactive key for use both on IBM compatible and Apple Macintosh personal computers. Printed and computer formats are supplied in one package and are intended to be complementary. CSIRO, Australia, Aus\$195, US\$195.

Fringe physics

Peter Holland

The Undivided Universe: An Ontological Interpretation of Quantum Theory. By D. Bohm and B. J. Hiley. Routledge: 1993. Pp. 397. £25, \$29.95.

WITH the death of David Bohm in 1992, twentieth-century theoretical physics lost one of its most important thinkers. In conventional physics Bohm is known for his contributions to plasma physics, his classic text on quantum theory and as codiscoverer of the Aharonov-Bohm effect. But his most enduring bestowal may well prove to be his work in an area many have felt to be on the fringes of acceptable physics: the interpretation of quantum mechanics that he first proposed in 1952 following the earlier work of Louis de Broglie. After years of neglect, the theory is undergoing a renaissance and this book presents what Hiley states in the preface to be Bohm's final thoughts on it.

Bohm's idea is that the electron is a point particle as in classical mechanics, but that its trajectory is influenced by what he called the quantum potential, a new type of non-classical physical field derived from the wave function. He thus provides an 'ontology', a model of how the world is built, that contrasts with the conventional interpretation which is concerned solely with our (statistical) knowledge of the world. In his original papers he showed how the statistical predictions of quantum mechanics could be generated through an

ensemble of individually well defined causal motions.

The book is intended to bring this idea up to date and describes in mathematical terms how to develop ontological explanations for many classic quantum phenomena, including transitions, interference and relativistic effects. A careful analysis is given to show that the causal theory cannot be excluded on grounds of naturalness or economy of concepts, for all the interpretations of quantum mechanics proposed so far entail assumptions that go beyond the basic formalism. But Bohm's theory has the major advantage of conceptual clarity, and the paradoxes that afflict other views simply do not arise.

One reason the theory has been dismissed is that it implies a deep interconnectedness of distant systems, or 'nonlocality', although, as pointed out by the authors, the orthodox view also involves nonlocality in an essential way. Indeed, the authors show a deeper appreciation of the meaning of the conventional interpretation than do many of its proponents. Most contemporary presentations still imply that history bequeathed us a homogeneous quantum world view. It is rarely remarked that Bohr, Heisenberg and von Neumann differed significantly in their analyses, von Neumann's concept of the 'quantum state' of an atomic system being incompatible with Bohr's notion that systems cannot be conceived independently of the means by which they are observed. It would help students considerably if this confusion in the official quantum story was openly admitted.

The authors attempt to find a metaphor

that adequately encapsulates what they feel is the novelty of quantum theory as revealed in this model. Their solution is to regard the quantum field as a repository of potentially active information, the electron being likened to a ship moving under its own energy and guided by a radar wave of considerably less energy. It is a view that some may find unpalatable because it seems to read more into the mathematics than is actually there. And there is a curious empiricist tone to some of the statements, such as the assertion that the only intrinsic property of an electron is its position because only this can be measured without disturbance. These ideas are interesting but there are other points of view consistent with the basic model that are not represented and the book does not provide a comprehensive review of the field as it now exists. This subject is beginning to flourish and the inclusion of other relevant recent work would have improved the presentation. There also seem to be some oversights in properly attributing previous works to their original sources.

The causal theory provides a simple criterion for judging when quantum systems will display classical behaviour (be governed by Newton's laws) through the relative effectiveness of the quantum potential. This suggests the view expressed in the title of the book that the Universe is an 'undivided whole', a system that is basically quantum mechanical yet capable of the classical behaviour we encounter at the scale of macro-experience. It is a compelling picture, but is it true that all the phenomena met at the macro-scale that are correctly described by classical physics are special cases in an enveloping quantum description? Although not mentioned in the book, it is a bonus of the causal model that it intimates a more subtle connection between the two domains, that classical phenomena cannot always be obtained in some limit, and it is unclear whether the Universe really is 'undivided' in the sense that the author intends.

Should we be alarmed that both causal and non-causal theories can account for the same empirical quantum data? Some may fear the lack of certainty in science, but it is closed minds that are harmful, not the proliferation of ideas. It was contemplation of Bohm's theory 30 years ago that prompted Bell to formulate his famous theorem distinguishing local from nonlocal theories. I hope that this book will further stimulate current investigations into causal theories in quantum physics. □

Peter Holland is in the Laboratoire de Gravitation et Cosmologie Relativistes, Université Pierre et Marie Curie, Boîte 142, 4 Place Jussieu, 75252 Paris Cedex 05, France.

The human IgE network

B. J. Sutton & H. J. Gould

The Randall Institute, King's College London, 26–29 Drury Lane, London WC2B 5RL, UK

IgE and its receptors are believed to have evolved as a mechanism to protect mammals against parasites. But other and intrinsically innocuous antigens can subvert this system to provoke an allergic response. For human populations in industrialized countries, allergy and asthma now represent a far greater threat than parasitic infection, and the main impetus for current studies of the IgE system is the hope of understanding and intervening in the aetiology of allergic diseases.

IMMUNOGLOBULIN E (IgE) occurs only in mammals and is one of five classes of antibody recognized in man. Its role in defence against parasites was first established by work on the cell-mediated killing of schistosomes *in vitro*^{1,2}, as well as by epidemiological studies in areas of endemic schistosomiasis and other parasitic diseases^{3,4}. IgE-secreting B cells are abundant in the skin, lungs and gut, the main sites of parasitic invasion. IgE elicits a range of cellular responses to antigens, culminating in anatomical and physiological changes which exclude parasites from the body. These include inflammation, itching, coughing, lacrimation, bronchoconstriction, mucus secretion, vomiting and diarrhoea, all common symptoms in allergic disorders.

Allergy is generally caused by the overproduction of IgE in response to common environmental antigens, such as those present in pollen, foods, house dust mites, animal danders, fungal spores and insect venoms. Predisposition to allergy, however, appears to result from an interaction between genetic and environmental factors. The most common allergic diseases are asthma, allergic rhinitis (hay fever), atopic dermatitis and food allergies, but the most dangerous is anaphylactic shock, usually provoked by insect stings or parenteral medication. Allergy in one form or another afflicts more than 20 per cent of the population, and the alarming increase in its prevalence, morbidity and mortality over the past decade has led to its designation as the "number one environmental disease"^{5,6}.

Existing drugs offer partial relief of symptoms in some of these conditions, but there is a clear need for more effective treatments. Recent advances in our understanding of the structure, function and synthesis of IgE and its receptors now promise new approaches to the design of therapies for allergy.

Changing perspectives in the IgE system

IgE constitutes a minuscule fraction of the total antibody in human serum (50–300 ng ml⁻¹, as compared with 10 mg ml⁻¹ of IgG) and is therefore of little help in neutralizing antigens. However, its action is powerfully amplified by the activities of the receptors to which it binds. The 'high-affinity' receptor on skin mast cells and blood basophils, FcεRI (receptor for the Fc region of the ε-chain of IgE), is responsible for immediate hypersensitivity reactions; when a multivalent allergen associates with FcεRI-bound IgE, it crosslinks receptor molecules on the cell membrane. This triggers degranulation of the cell, rapid release of stored mediators, notably histamine, and the synthesis and secretion of cytokines that attract and activate inflammatory cells⁷. Although FcεRI has recently been discovered on Langerhans cells^{8,9}, eosinophils^{10,2} and activated monocytes (G. Stingl, personal communication), its biological functions in these cell types have not been identified, except that in eosinophils it can mediate IgE-dependent killing of schistosomes *in vitro*^{10,2}.

Activation of inflammatory cells by the products of mast-cell degranulation induces a 'low-affinity' IgE receptor, FcεRII, also known as CD23. This receptor was previously thought to be responsible for the IgE-dependent cytotoxic activities of inflammatory cells, and it does indeed mediate the phagocytosis of immune complexes by monocytes *in vitro*^{10,11}. The recent observation that FcεRI as well as FcεRII is expressed on eosinophils^{12,10,2}, monocytes and Langerhans cells^{8,9,13}, however, raises the question of which receptor mediates which effector function of IgE.

The role of FcεRII in the humoral immune response has only recently come to light (see refs 14–16 for reviews). On the surface of B cells, it plays a part in IgE-dependent antigen presentation to T cells and in adhesion of B cells to each other. It is also expressed copiously on follicular dendritic cells (FDC) and may thus be involved in the homing of B cells to the germinal centres of secondary follicles in the lymph nodes and spleen. Soluble fragments of FcεRII (sFcεRII) stimulate the growth and differentiation of the precursors of plasma cells, T cells and basophils, thereby initiating both humoral and cell-mediated immune responses. Although FcεRII is C-type lectin, it does not bind to the carbohydrate component of IgE; the inference that this part of the molecule must recognize something else has recently

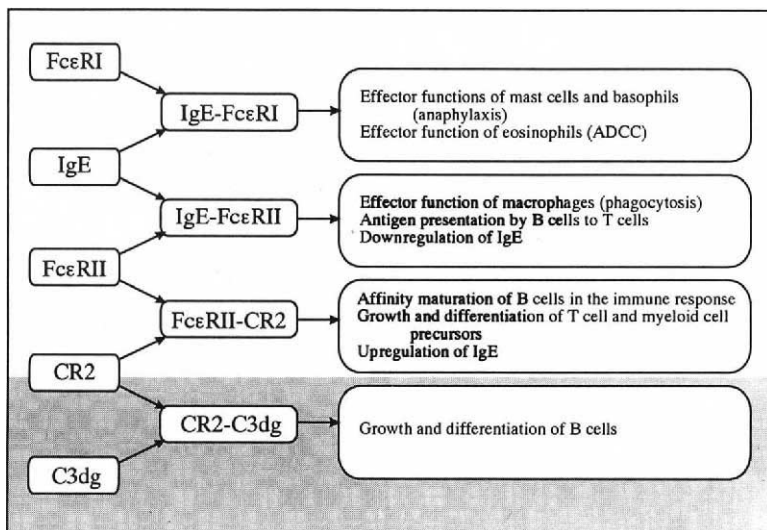


FIG. 1 Principal components of the IgE network. IgE, FcεRII and CR2 each have two or more natural ligands, and the physiological functions of the different complexes, and the cells in which these occur, are listed. The shaded area represents the overlap of the IgE antibody and complement systems; C3dg is one of the active fragments of the complement component C3.

been borne out with the discovery that it binds to CR2 (complement receptor 2)¹⁷. This protein, also called CD21, was already known to be a receptor for fragments of complement component C3 and the Epstein-Barr virus (EBV). It appears to mediate two important cytokine activities ascribed to FcεRII, the upregulation of IgE synthesis¹⁷ and the rescue of germinal centre B cells from apoptosis¹⁸. IgE itself also closes a negative feedback loop by switching off its own synthesis, through interaction with FcεRII.

These observations can now be interpreted in terms of a system of interacting proteins which we call the IgE network; Fig. 1 presents the principal components of this network and summarizes the physiological functions triggered by their interactions.

The IgE molecule

Prausnitz and Küstner¹⁰³ first demonstrated the presence of a factor in the blood of allergic subjects which, when transferred to the skin of non-allergic individuals, rendered them sensitive to allergens. This substance was identified as a cytophilic antibody by the Ishizakas¹⁰⁴, but its concentration was below the threshold of detection by protein assay. It could, however, be separated from the known immunoglobulins, IgM, IgG and IgA, and was named IgE from the erythema that the allergens provoke in sensitive skin¹⁹. The first tangible evidence that IgE existed came from a rare myeloma, and the protein from this cell line afforded a standard for the measurement of IgE concentrations in blood, as well as material for the first structural studies. The cells were also the source of messenger RNA for cloning the ε-chain cDNA, the starting point for some of the later structural studies of IgE (reviewed in ref. 20).

The protein sequence revealed that IgE shares the overall structure of other classes of antibody, but is distinguished from them by the sequence of its (ε) heavy-chain constant region. Like the μ-chain of IgM, this consists of four constant domains (Cε1–Cε4), one more than in IgG, IgA or IgD. The additional domain (Cε2) replaces the hinge region of the other antibody classes (Fig. 2). The two C-terminal domains of IgE (Cε3 and Cε4) are highly homologous to those of IgG, and can be modelled on the basis of the crystal structure of human IgG Fc (Fig. 2; refs 21, 22).

Limited proteolysis of IgE yields an Fc fragment (Cε2–Cε4) that can still bind FcεRI and FcεRII but isolated Cε2 and Cε4 domains cannot. Synthesis of recombinant peptides in *Escherichia coli* or mammalian cells, and the exchange of domains between rodent and human IgE have provided direct evidence that the FcεRI binding site is located in Cε3, and it is now clear that the first twelve amino acids of Cε3 are necessary for binding (see refs 20, 23 for reviews), although other residues in Cε3 or Cε4 can perturb the interaction^{24,25}. The solvent-accessible part of this dodecapeptide sequence occupies a position near the central axis of the Cε3 domain pair in the predicted structure of IgE (Fig. 2). Although monomeric recombinant peptides are recognized by FcεRI, only one of the two sites appears to be accessible in the native IgE molecule.

The binding site for FcεRII has been mapped to Cε3 by similar means^{24,26}, but deletion of the first twelve amino acids of the domain does not eliminate binding, and it must therefore be distinct from the FcεRI site. More precise mapping is needed, but antibody blocking and peptide inhibition experiments²⁶ already indicate that the site is near the glycosylated residue Asn 371 (Fig. 2). If so, the two FcεRII binding sites in IgE are as far as possible from each other and from the central axis of the molecule. The receptor binds only to disulphide-linked ε-chain fragment dimers, and is probably itself at least a dimer able to bind both sites on IgE at once (compare Fig. 6a). Both FcεRI and FcεRII bind to IgE peptides synthesized in *E. coli*^{20,26}, indicating that the carbohydrate components of IgE (Fig. 2) are not required for either interaction. (A soluble S-type lectin that can bind to the carbohydrate components, known as Mac-2 or εBP, has recently been found on human neutrophils²⁷; it may thus

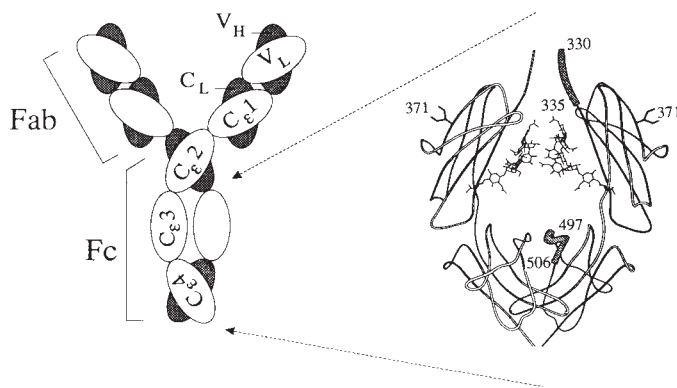


FIG. 2 The structure of human IgE. The IgE molecule consists of two identical light (L) chains (κ or λ) and two identical ϵ heavy (H) chains, folded into domains as shown on the left. Also indicated are the two antigen-binding fragments (Fab), and the constant (C) region fragment (Fc) in which the receptor binding sites are located. In each Fab, a disulphide bridge links CL to Cε1, and in Fc, two bridges link the Cε2 domains of the ε-chains. On the right, the predicted fold of the polypeptide chain within the Cε3 and Cε4 domains is shown, together with the complex-type oligosaccharide chains that are N-linked to Asn 394 and lie between the Cε3 domains. This glycosylation site is conserved in other immunoglobulin classes; a second ε-specific glycosylation site at Asn 371 is also indicated schematically. The structure of these domains has been modelled^{21,22} on the homologous crystal structure of IgG Fc. A segment of the ε-chain sequence (residues 330–335) that has been implicated in FcεRI binding is highlighted. This region, at the N-terminal end of the Cε3 domain (and shown for only one ε-chain), is expected to be accessible, although its conformation is uncertain and may not be as extended as shown here. The region indicated in Cε4 (residues 497–506) corresponds to the peptide that has been proposed as a vaccine for the treatment of allergy (see text, and ref. 87).

play a part in parasitic infections and/or allergy¹⁰⁵, but will not be discussed further here as little is known about its structure at present.)

One striking feature of IgE that has only recently become apparent is that, in contrast to IgG, it is bent. Resonance energy transfer between fluorophores attached to different parts of the chain reveals that the distance between the N- and C-termini of the IgE molecule in solution is no more than 7 nm, compared with the 17.5 nm implied by a planar structure²⁸. The shape of IgE does not appear to change when it binds to FcεRI, and distances between the two fluorophores and others dispersed in the membrane indicate that the Fc region of a receptor-bound IgE molecule lies parallel to the membrane surface²⁸. The discovery that IgE is bent (as shown in Figs 4 and 6a) profoundly alters our view of its interactions with FcεRI and FcεRII, providing a new framework for the interpretation of existing experimental data.

The high-affinity receptor FcεRI

The rat basophilic leukaemia (RBL) cell line was the first source of FcεRI, and of information about its interaction with rodent IgE and the way in which it activated cells. No comparable human cell line exists even today, although the genes for the human receptor have been expressed in non-basophilic human cell lines and in the RBL cell line^{29–31} to allow studies of the interaction with human IgE (see refs 32, 33 for reviews).

FcεRI consists of four transmembrane polypeptides with the composition $\alpha\beta\gamma_2$, and although it is the α-chain that binds IgE, the β- and γ-chains are required for insertion of the α-chain into the membrane and for signal transduction. The extracellular portion of the α-chain contains two domains characteristic of the immunoglobulin superfamily (Fig. 3); it thus resembles the FcγR, FcαR, and poly-Ig receptors, which contain two or more extracellular immunoglobulin-like domains³². The extracellular

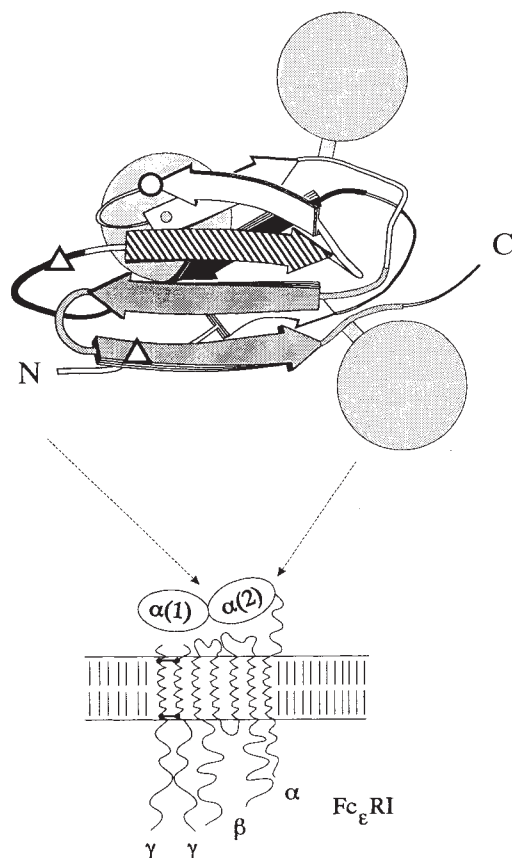


FIG. 3 Predicted structure of Fc ϵ RI α -chain. Top, predicted fold of the polypeptide chain within the second of the two immunoglobulin-like extracellular domains, $\alpha(2)$, modelled on homologous regions of known immunoglobulin superfamily domain structures, in particular CD2 (ref. 100). Strands of β -sheet are represented by arrows; the single intradomain disulphide bridge is not shown. Bottom, schematic representation of the IgE binding α -chain, β - and two disulphide-linked γ -chains of Fc ϵ RI. In the model of the $\alpha(2)$ domain, three segments are highlighted: black, peptide bound by an antibody that inhibits IgE-Fc ϵ RI interaction³⁸; cross-hatched, sequence of 7 residues conserved between the rat and human sequences and proposed as part of the IgE binding site¹⁰¹; stippled, segment corresponding to the homologous sequence of Fc γ RII that confers IgG binding when transplanted into Fc ϵ RI (refs 36, 37). Allelic variants of Fc γ RII have also been identified that affect IgG binding, and the homologous locations of these residues are indicated on the model: circle, residue 131; triangles, residues 116 and 161 (reviewed in ref. 36). The three glycosylation sites in $\alpha(2)$ are represented by the large shaded circles, which nevertheless underestimate the volumes occupied by complex-type N-linked oligosaccharide chains. The locations of all of these features are consistent with homologous Ig-binding regions in Fc ϵ RI and Fc γ RII that comprise residues of the four-stranded face of the domain, and adjacent loops.

region of Fc ϵ RI α -chain most closely resembles that of Fc γ RII (CD32) and Fc γ RIII (CD16), and indeed the latter has been shown to form a complex with the β - and γ -chains of Fc ϵ RI (refs 32, 34). Fc ϵ RI(α) is heavily N-glycosylated, but the carbohydrate component is not required for IgE binding¹⁰⁶.

The immunoglobulin-like domains of the receptor α -chain, although very similar to antibody V and C domains, form a recognizable subgroup within the immunoglobulin superfamily, which has been termed C2 (ref. 35). The atomic structure of these domains is still unknown, but their homology with known immunoglobulin-like domains allows them to be modelled with some assurance (Fig. 3). Studies of human Fc ϵ RI(α)-Fc γ RII (refs 36, 37) and interspecies Fc ϵ RI(α) chimaeras^{106,107} place the IgE binding region in the second domain ($\alpha(2)$ in Fig. 3), in accord with earlier conclusions based on the location of an epi-

tope recognized by an inhibitory antibody³⁸. A more precise definition of the site will require site-specific mutants, but mapping studies of the homologous interaction between Fc γ RII and IgG (Fig. 3) point to an exposed region of the four stranded face of the $\alpha(2)$ domain, and binding to this site would not be affected by the three putative glycosylation sites in the $\alpha(2)$ domain.

The interaction between IgE and Fc ϵ RI(α) may thus be envisaged as shown in Fig. 4. The IgE molecule binds with its convex surface facing the membrane²⁸. The receptor binding site, located in the N-terminal segment of C ϵ 3 that adjoins C ϵ 2 (Fig. 2), is thus accessible only on the same convex face; the second site will be occluded on the concave surface. This is consistent both with the presence of a single α -chain in the receptor, and its affinity for single ϵ -chains. It also explains why only one Fc ϵ RI α -chain can bind to IgE^{39,106}. The configuration of Fc ϵ RI(α) in the complex may depend on its interaction with IgE, as the kinetics of association have been taken to imply a conformational change upon IgE binding⁴⁰.

The low-affinity receptor Fc ϵ RII

Fc ϵ RII was independently characterized as a B-cell receptor for IgE, a marker for activated B cells (named CD23), and a marker for EBV-transformed B cells (named EBV-CS2). It was later found on various inflammatory cells (macrophages, eosinophils, platelets, natural killer (NK) cells), T cells, FDCs, Langerhans cells and epithelial cells of bone marrow and thymus^{14,15}, and is the only known antibody receptor that is not a member of the immunoglobulin superfamily. The C-type (calcium-dependent) lectin domain in its extracellular sequence places it in a family of proteins that includes the asialoglycoprotein receptor (ASGPR) and the adhesion proteins known as 'selectins'⁴¹. In common with the former, Fc ϵ RII has a cytoplasmic N-terminus and thus is classified as a type-II integral membrane protein.

This lectin domain is flanked on the C-terminal side by a sequence of unknown tertiary structure containing an inverse RGD (Arg-Gly-Asp) sequence which may act in cell adhesion¹². Although two ligands for Fc ϵ RII are already known, IgE and CR2, the presence of this adhesion motif implies that there may be a third, perhaps an integrin. On the N-terminal side the lectin domain is flanked by a sequence containing heptad hydrophobic repeats, which is predicted to form an α -helical coiled-coil 'stalk'⁴² and mediates trimer formation⁴³; this is followed by the

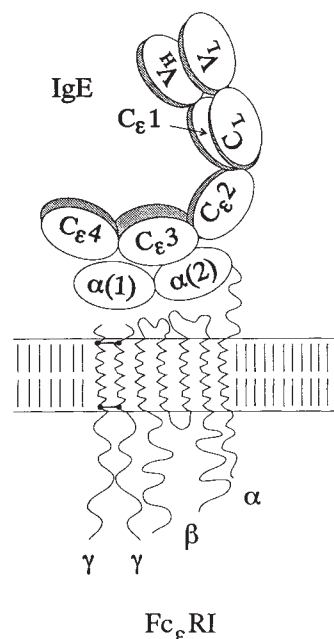
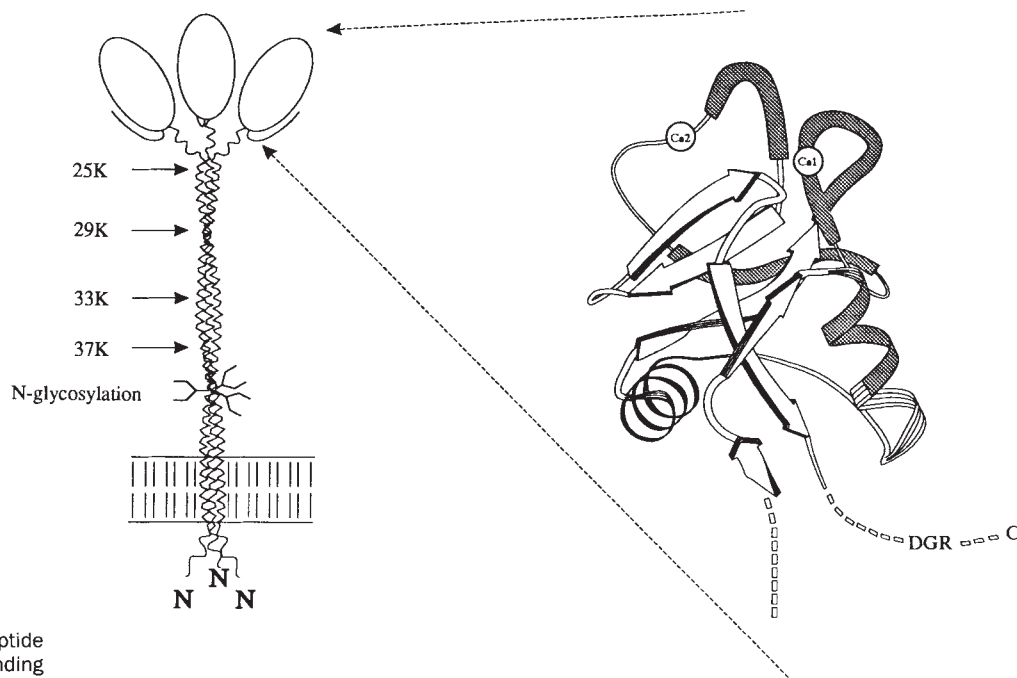


FIG. 4 Schematic representation of an IgE molecule bound to Fc ϵ RI. The IgE molecule adopts a bent conformation consistent with spectroscopic data²⁸, and the N-terminal region of the C ϵ 3 domain of IgE (compare Fig. 2) contacts the second domain of Fc ϵ RI α -chain (compare Fig. 3). The orientation of IgE relative to the membrane is also in accordance with experimental data²⁸.

FIG. 5 Schematic representation of human FcεRII, showing the triple α-helical coiled-coil stalk region (length ~15 nm), and the three heads that contain the C-type lectin homology regions. The single glycosylation site in each chain is shown schematically (not to scale) and the sites of proteolysis in the stalk region are indicated, together with the sizes of the soluble fragments released. The structure of the lectin-like head region shown on the right is based on the crystal structure of the homologous region of the mannose-binding protein⁴⁶. Of the two Ca²⁺ binding sites of MBP, only Ca2 is conserved in human FcεRII, although both appear to be conserved in mouse FcεRII.

The four stippled segments of polypeptide chain are those implicated in IgE binding in studies with chimaeric proteins⁴⁷; the three loop segments form a contiguous surface region, the location of which is consistent with the Ca²⁺ dependence of IgE binding. The C-terminal extension to the lectin homology region that in human FcεRII



contains the reverse RGD sequence is indicated. (The four disulphide bridges within this region of human FcεRII are not shown.)

transmembrane sequence and a short cytoplasmic segment (Fig. 5). There are two forms of FcεRII, FcεRIIa and FcεRIIb, differing only in their 6/7 N-terminal amino acids¹⁵. They are derived from the same gene by the use of different promoters (a and b) controlling separate first exons, which are spliced to a common mRNA sequence. FcεRIIa is a developmentally regulated gene, expressed only in antigen-activated B cells before their differentiation into immunoglobulin-secreting plasma cells. FcεRIIb expression is inducible by interleukin-4 (IL-4) on all of the cell types mentioned so far.

The membrane form of FcεRII gives rise to soluble fragments (sFcεRII) by proteolysis. The initial cleavage, within the 'stalk' and close to both the cell membrane and the single site of N-glycosylation, releases a fragment of M_r 37,000 (37K) (Fig. 5). The nature of the proteases that cleave FcεRII *in vivo* is of considerable interest, as is the suggestion that the first to act may be, or require, sFcεRII itself¹⁵. The 37K fragment can be cleaved nearer the lectin domain to give fragments of 33, 29 and 25K containing the lectin domain and the C-terminal 'tail', and a 16K fragment lacking the tail (Fig. 5). It is remarkable that the fragments of different sizes appear to have quite different activities. All retain binding specificity for IgE, although the smaller fragments bind with lower affinity than the intact molecule. Only fragments of $M_r \geq 29K$ promote IgE synthesis, however, and indeed the 16K fragment inhibits this activity⁴⁴. The remaining cytokine activities are common to the intact molecule and all the fragments^{15,45}.

The recently determined crystal structure of the mannose-binding protein (MBP; ref. 46) makes it possible to model the structure of the lectin domain of FcεRII as shown in Fig. 5. The regions implicated in IgE binding⁴⁷ (shown stippled in Fig. 5) lie close to the Ca²⁺ binding site which is conserved in FcεRII (Ca2; Fig. 5) and forms part of the mannose-binding site in MBP⁴⁸, explaining why IgE binding is both Ca²⁺-dependent⁴⁹ and inhibited by fucose-1-phosphate¹⁵, even though binding to IgE does not involve carbohydrate²⁶. Fucose is implicated in the binding of FcεRII to CR2 (ref. 50), however, and if Ca2 indeed marks the carbohydrate binding site in FcεRII, the IgE- and CR2-binding sites must overlap (as implied by experimental data⁴⁵), so that IgE would compete with CR2 for binding to

FcεRII (see later). The model of FcεRII (Fig. 5) further predicts that the ligand-binding sites are remote from both the stalk and the Asp-Gly-Arg (DGR)-containing 'tail'.

The release of soluble fragments from membrane-bound FcεRII is inhibited by IgE⁵¹ and so IgE and IgE-antigen complexes repress IgE synthesis⁵². A possible mode of interaction between IgE and FcεRII is shown in Fig. 6a, with two lectin heads binding to the two sites on the IgE molecule, but it is clear from the relative dimensions of the two molecules that IgE cannot exert a direct steric effect upon the primary site of proteolysis (which releases the 37K fragment; Fig. 5). We have therefore suggested an allosteric mechanism⁵³ by which IgE binding stabilizes the coiled-coil stalk and protects it against proteolysis. This effect is mimicked by most monoclonal antibodies against FcεRII, but a small group of antibodies exert the opposite effect and promote IgE synthesis. This surprising observation, together with the fact that Fab fragments operate as effectively as the intact antibodies, cannot be explained in terms of receptor crosslinking, prompting consideration of an allosteric mechanism⁵⁴. We can now rationalize the effect of this second group of antibodies and Fab fragments in terms of a destabilization of the coiled-coil stalk and enhanced release of sFcεRII.

B-cell FcεRIIa has another function, as antigen-IgE-FcεRII complexes present antigen very effectively to T cells⁵⁵. Presentation entails internalization, degradation within a cytoplasmic compartment, and recycling of fragments to the cell surface. The role of FcεRII in endocytosis is thus similar to that of the homologous ASGPR in the degradation of glycoproteins by the liver. ASGPR is a trimer with a predicted structure resembling FcεRII (ref. 42); its activity depends upon its oligomeric character, which allows it to be crosslinked by multivalent ligands⁵⁶. By analogy, FcεRII's capacity for antigen presentation probably depends on its oligomeric structure too.

The trimeric structure of FcεRII may also account for the contrasting cytokine activities of the different soluble fragments. Only fragments larger than 25K, which retain part of the stalk, should remain as trimers. Such fragments would be able to crosslink membrane-bound IgE and CR2 on the B-cell surface, allowing IgE class selection by 37K and 29K (but not 25K)

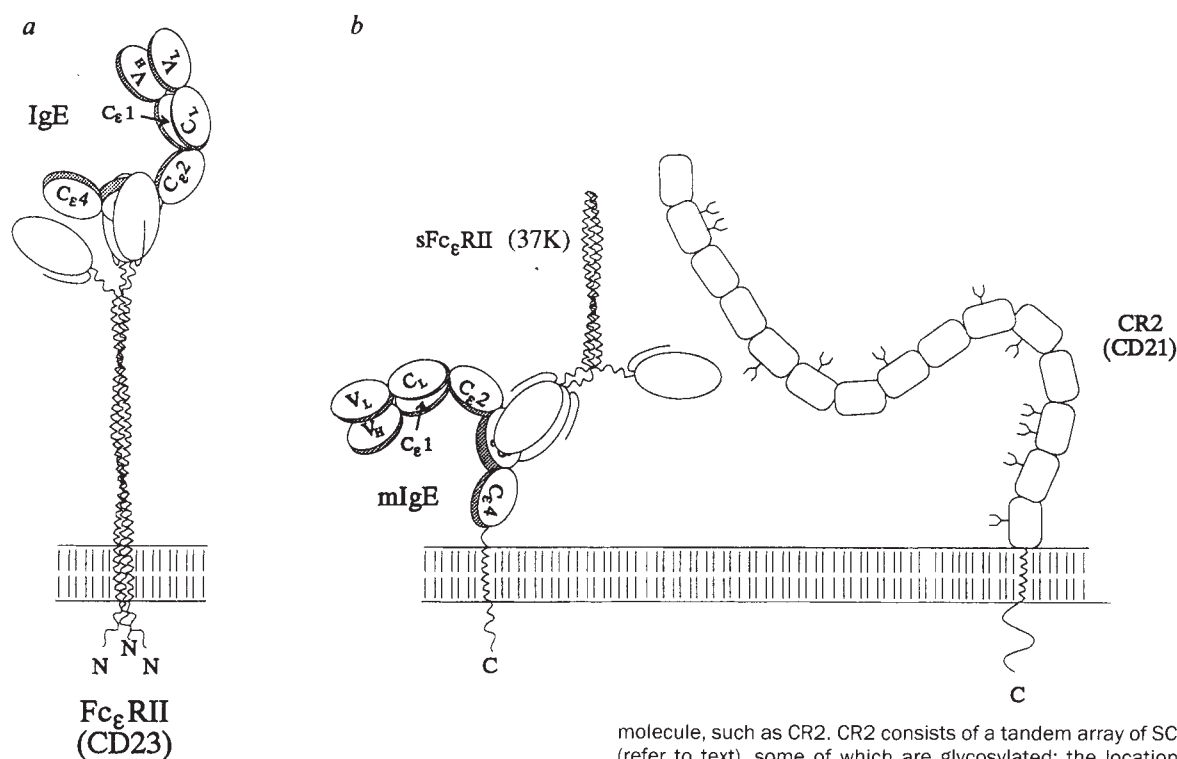


FIG. 6 Multiple interactions between IgE, FcεRII and CR2. *a*, The interaction between human FcεRII and IgE. Two lectin-like head regions of membrane-bound FcεRII combine with the two Cε3 domains of IgE. *b*, Interaction of trimeric 37K sFcεRII with membrane-bound IgE. The 'free' head may bind to another IgE molecule, or other cell surface

molecule, such as CR2. CR2 consists of a tandem array of SCR domains (refer to text), some of which are glycosylated; the locations of these sites are shown schematically. FcεRII may bind in the region of domain 5 (refer to text) and involve the recognition of carbohydrate. Crosslinking mIgE and CR2 (which forms a signal transduction complex with CD19; see text and Fig. 7) on the surface of a committed B cell may provide a mechanism for specific activation of IgE synthesis by FcεRII. (Molecules and cell membrane are drawn to scale.)

sFcεRII, as depicted in Fig. 6*b*. The transient nature of the fragments also suggests a mechanism for localizing their activity.

CR2 as counter-receptor for FcεRII

The discovery that CR2 is a counter-receptor for FcεRII resolved several questions at once, for it established that CR2 is the ligand for the inferred lectin function of FcεRII (refs 17, 50), pointed to CR2 as the missing receptor required for cell signalling by FcεRII, and explained the link between FcεRII expression and B-cell transformation by EBV⁵⁷. Fragments iC3b, C3dg and C3d of the complement component C3, the EBV coat protein gp350/220 and α -interferon had previously been identified as ligands for CR2, and all of these molecules can deliver growth signals to B cells on binding (reviewed in ref. 58). CR2 was therefore thought to link the complement and antibody immune systems; it now appears to link the humoral and cellular arms of the latter as well.

CR2 is a highly glycosylated membrane protein found on B-cells, FDCs and some T cells and basophils. Its extracellular region comprises 15 or 16 homologous units in tandem, each made up of 60–75 residues, with the compact fold of the short consensus repeat (SCR; ref. 59). This produces an extended flexible structure some 40 nm long⁶⁰, as shown in Fig. 6*b*, which also indicates the potential *N*-glycosylation sites. Several species of CR2 can be separated electrophoretically, only one of which is recognized by FcεRII (ref. 17); these may differ solely in their glycosylation. Differential glycosylation of CD21 is also of interest in relation to the recently reported ability of CD23 to induce degranulation of basophils by the crosslinking of CD21 (ref. 112). The binding sites for C3 fragments and EBV have been mapped with monoclonal antibodies to the two N-terminal domains^{61,62} and do not involve carbohydrate. By contrast, FcεRII binding is inhibited by carbohydrate⁵⁰ and by monoclonal antibodies¹⁷ with epitopes that overlap in domain 5 (refs 18, 61). The valency of the ligand is important for both the C3

fragments and EBV gp350/220, because B-cell proliferation is promoted by multivalent but inhibited by univalent ligands^{63,64}. It remains to be seen whether the stimulatory and inhibitory activities of different sFcεRII fragments can be rationalized in the same way.

sFcεRII can trigger CR2 on B-cells and enhance IgE synthesis (ref. 17; see also Fig. 6*b*), but the two molecules may elsewhere engage in adhesive interactions. FcεRII expression on the surface of B cells induces the homotypic adhesion of B cells, and sFcεRII inhibits the migration of monocytes (reviewed in ref. 16). In this case, the role of FcεRII is reminiscent of that of another group of C-type lectins, the selectins⁴¹. Most recently, the FcεRII-CR2 interaction has been implicated in the rescue of germinal-centre B-cells from apoptosis¹⁸; it thus participates in perhaps the most fundamental event in the immune response, the clonal amplification of B-cells selected by antigen. The possible contribution of the FcεRII-CR2 pair to this process is depicted in Fig. 7. Immune complexes held by FDCs can interact with membrane IgM (mIgM) on the surface of B cells, but this interaction alone is insufficient to stimulate the B cells because both antigen concentration and affinity are too low. Co-ligation of the B-cell protein CD19 can bring the threshold for activation into the physiological range⁶⁵, and the activation of mIgM and CD19 is synergistic⁶⁶. A ligand for CD19 has been sought without success, but CD19 forms a membrane complex with CR2 (ref. 67) and FcεRII can therefore act as ligand for this complex. The presence of C3 fragments bound to the immune complex may also serve to crosslink the two membrane complexes (Fig. 7) contributing to the observed synergism.

The demonstration that CR2 mediates the cytokine activities of FcεRII offers a new perspective on the process of B-cell transformation by EBV. Not only does EBV use CR2 to gain entry, but once inside it expresses two proteins, EBNA-2 and LMP-1, which activate the proto-oncogene *bcl-2* (implicated in the mechanism of cell survival⁶⁸), as well as the genes for FcεRII

and CR2 (ref. 69). Although FcεRII expression is normally shut down in the terminally differentiated cell, gene activation by EBNA-2 and LMP-1 is permanent, and hence infected cells continue to express both growth factor (sFcεRII) and receptor (CR2) indefinitely. Incubation of germinal-centre B cells with sFcεRII and IL-1 *in vitro* leads to expression of *bcl-2* (ref. 70), and FcεRII may therefore mediate the upregulation of *bcl-2* by LMP-1. CR2 can also bind the anti-oncoprotein p53 in transformed cells⁷¹; FcεRII and CR2 are thus clearly implicated at several stages in the mechanism of transformation of B cells by EBV.

Receptor coupling to intracellular signalling

Each of the four proteins that we have discussed exists in one or more membrane-bound forms acting as signal receptors. Propagation of these signals depends on the interactions between the receptor and other membrane-bound and cytoplasmic proteins. These interactions are by no means as well defined as the complexes formed by the receptors with each other, but some essential sequence motifs, and the proteins that recognize them, have been identified.

Membrane-bound IgE is expressed on cells committed to IgE synthesis and serves as an antigen receptor. As with other membrane immunoglobulins, it forms a complex with two disulphide-linked heterodimeric proteins, made up of immunoglobulin α- and β-chains, which function in signal transduction (see ref. 72 for review; see mIgM in Fig. 7). The cytoplasmic sequences of Ig-α and Ig-β each contain a 26-amino-acid sequence motif, (D/E)X₇(D/E)X₂YX₂(L/I)X₇YX₂(L/I), referred to as the antigen-receptor homology-1 (ARH1) motif, which associates with tyrosine kinases of the *src* family and phosphatidylinositol 3-OH kinase (PtdIns-3 kinase)⁷³. Similarly, antigen signalling by the IgE-FcεRI complex is mediated not by the ligand-binding α-chains, but by the associated β- and γ-chains, which also contain the ARH1 motif and associate with *src*-like kinases⁷⁴. Phosphorylation of residues within these motifs and in the associated proteins occurs during the early steps of signal transduction and modulates these interactions.

The short cytoplasmic region of FcεRII does not contain an ARH1 motif, and the mechanism of signal transduction by this receptor is unknown (though a *src*-like tyrosine kinase has been found associated with FcεRII in a human NK-like cell line⁷⁵). The a form of FcεRII, however, does contain the sequence YSEI, an example of the tetrapeptide motif that acts as a signal for transport to endosomes via coated pits⁷⁶. Thus FcεRIIa, but not b (which lacks this motif), can function in antigen presentation; conversely, the dipeptide motif NP present in the cytoplasmic sequence of FcεRIIb, but not a, is required for this receptor to mediate phagocytosis in monocytes¹¹. As the sequence YX₂(L/I) occurs (twice) within the ARH1 motif, its presence in the β- and γ-chains of FcεRI may be responsible for internalization of this receptor also.

As mentioned above, CD19 forms a complex in the membrane with CR2, and its extensive cytoplasmic sequence contains two copies of the motif YX₂M, which has been characterized as a binding site for PtdIns-3 kinase^{77,108}; the two tyrosines are required for this association¹⁰⁸. Subsequent steps in signalling by this complex, and by the mIgM complex with which it exhibits synergy, have yet to be elucidated in detail, but the two pathways appear to be distinct⁶⁶.

Feedback control of IgE synthesis

Expression of IgE is controlled both by heavy-chain switching and by the selection of IgE-committed cells for survival and differentiation. The requirements for selective switching to IgE have been defined by studies of peripheral blood mononuclear cells *in vitro*^{14,15,78}, and three molecular signals appear to be involved. The first is IL-4, which directs recombination to the constant region of the ε-heavy chain. The other two are CD40

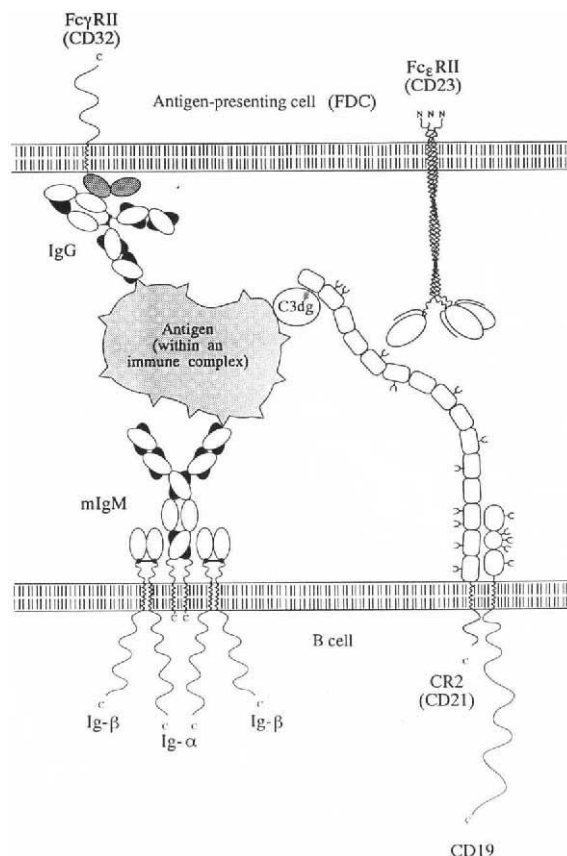
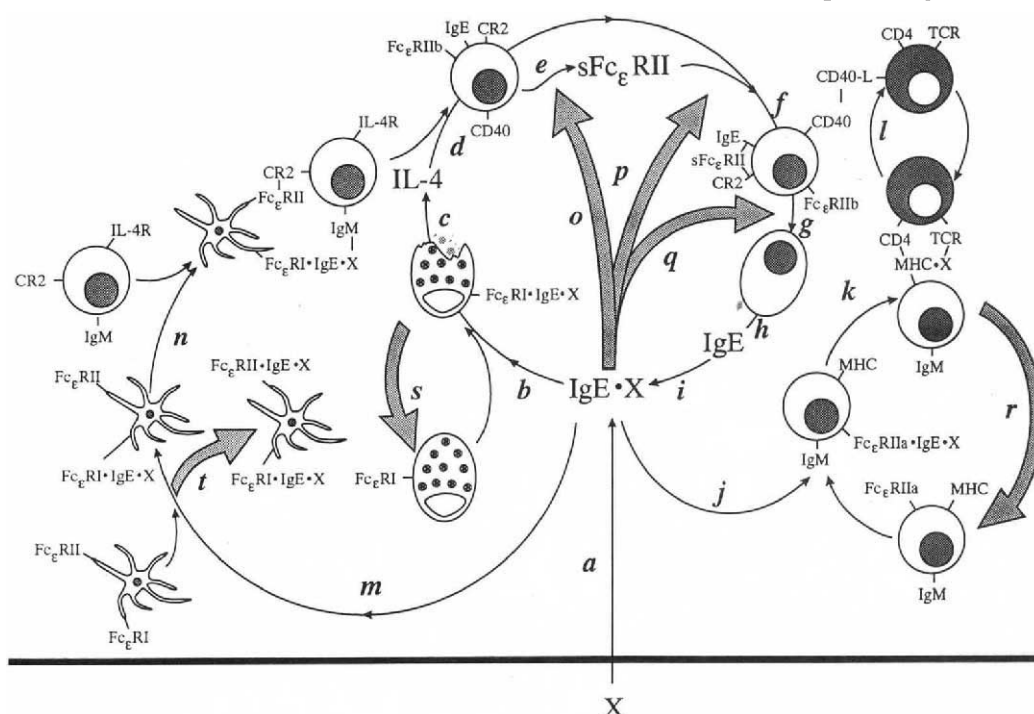


FIG. 7 Cell-cell interactions mediated by FcεRII and CR2. On the left, an IgG molecule attached to an antigen represents an immune complex that is bound to the surface of an antigen-presenting cell (such as a follicular dendritic cell, FDC) by means of an IgG Fc receptor (FcγRII is shown here). The immune complex is also bound to (and crosslinks) the B-cell antigen-receptor complex, which consists of mIgM and its associated disulphide-linked pairs of immunoglobulin α- and β-chains. These latter chains mediate signal transduction. On the right, FcεRII on the FDC interacts with CR2, which is in turn associated with CD19 in the B-cell membrane (and two other proteins, TAPA-1 and Leu-13; not shown). (CD19 consists of two extracellular Ig-like domains separated by a further truncated domain, and an extensive intracellular region.) The FcεRII–CR2 interaction promotes cell contact and delivers a signal to the B cell via CD19, which interacts synergistically with the signal transduced by mIgM and lowers the threshold for activation of B cells by immune complexes (see text). If complement activation also occurs, CR2 may interact with antigen-bound fragments of C3 (C3dg) in the immune complex, thereby enhancing the association between these two membrane complexes; the binding sites for C3dg (in domains 1–2) and FcεRII are distinct. A parallel mechanism might account for the presentation of IgE immune complexes by FcεRI on Langerhans cells which, like FDCs, also express FcεRII. The adhesion interaction between FcεRII and CR2 may also serve to enhance contact between other cells that express these molecules (B cells and T cells for example¹¹¹). Furthermore, the DGR-containing C-terminal extensions to each lectin domain of FcεRII may provide yet further opportunities for adhesive interactions or crosslinking of cell-surface molecules and generation of intracellular signals. (Molecules and cell membrane are drawn to scale.)

ligand (CD40-L) and FcεRII which induce the IgE-producing cells to become memory cells or plasmacytes respectively^{79,80}.

These choices are made, it is believed, in the germinal centres of secondary follicles in the lymph nodes and spleen; T cells are envisaged to be the source of IL-4 and to stimulate B cells through CD40-L, whereas FDCs are responsible for antigen presentation and the stimulation of B cells through FcεRII. It is now clear, however, that mast cells both secrete IL-4 (ref. 7) and express CD40-L, allowing them to perform the same activities in the skin, lungs and gut as do T cells⁸¹. IL-4 upregulates FcεRII on B cells and inflammatory cells, providing in turn a

FIG. 8 Feedback control of IgE synthesis. Positive signalling (thin arrows): a, Antigen X enters the system and forms a complex IgE.X; b, j and m are alternative fates for IgE.X at low concentrations of IgE. b, Binding of IgE.X to FcεRI on a mast cell. c, Secretion of IL-4 (and other inflammatory mediators that are not shown), activated by binding of IgE.X to FcεRI. d, Induction of FcεRIIb and CD40 expression, and class switching to IgE, in an antigen-activated B-cell by IL-4. e, Release of sFcεRII from the B cell itself or associated Langerhans cell. f, Binding of sFcεRII to CR2 and mIgE (Fig. 6b), promoting survival of the IgE-committed B cell. g, Differentiation of the cell into a plasmacyte, stimulated by the ligand of CD40 (CD40-L). h, Secretion of IgE. i, Binding of IgE to the antigen X. j, Binding of IgE.X to FcεRIIa on the surface of a B cell. k, Antigen presentation to a T-cell specific for the processed antigen X. l, Induced expression of CD40-L by the activated T-cell, and binding to CD40 on an IgE-committed B cell. Expression of CD40-L by mast cells (not shown) as well as T cells may obviate a strict requirement for the latter cells in the periphery. m, Binding of IgE.X to FcεRI on a Langerhans cell. n, Antigen presentation to a B cell involving the suggested co-ligation of FcεRI-IgM and FcεRII-CR2. Negative signalling (thick arrows): o, Inhibition of the cleavage of FcεRIIb



on the B cell (or Langerhans cell) by IgE.X. p, Inhibition of the binding of sFcεRII to CR2 on the B cell by IgE.X. q, Inhibition of the secretion of IgE, caused by IgE.X crosslinking of FcεRIIb on the B cell. r, Degradation of IgE in the course of antigen presentation, and s, after internalization by mast cells. t, Switching off the Langerhans cell by saturation of FcεRII at high IgE concentration.

local source of contact stimulation and of soluble growth factor. The co-expression of FcεRI and FcεRII in Langerhans cells^{8,9,13} may moreover imply that they could serve a function in the periphery analogous to that of FDCs in the germinal centres. These suggested interactions are summarized in Fig. 8.

Homeostasis also requires mechanisms that limit IgE production. Clearly IgE itself can exert negative-feedback control as its binding site on FcεRII overlaps with the site for CR2 (see above). IgE-antigen complexes are even more effective⁵² by reason of their higher affinity, and appear to act by blocking both the release of FcεRII (ref. 51) and the binding of sFcεRII to CR2. These feedback circuits will come into play only at IgE-antigen concentrations at which the sites on FcεRII are close to saturation. As Langerhans cells express both FcεRI and FcεRII, antigen presentation by these cells may be possible only when the former are occupied but the latter are free for interaction with CR2 (compare Fig. 7), thus being restricted to low IgE concentrations. In addition, IgE complexes can downregulate IgE synthesis by crosslinking FcεRII on B cells⁸² and the degradation of IgE in the course of antigen presentation mediated by FcεRII may also contribute to feedback control⁸³. All of these pathways for downregulation of IgE are represented in Fig. 8. The ability of FcεRII to promote germinal-centre B-cell survival, thymocyte maturation and myeloid cell differentiation through its interaction with CR2 suggests that it may also have a more general function in immunosuppression.

Prospects for new therapeutic strategies

The central position of IgE in the network of reactions leading to allergy commends it as a target for therapeutic intervention, whether at the level of its interaction with FcεRI, on the ensuing signalling pathways or the control of its biosynthesis.

Recombinant peptides comprising structural elements from IgE or FcεRI have been investigated as competitive inhibitors of the IgE-FcεRI interaction^{20,84,85}. The high affinity of IgE for

FcεRI ($K_a = 10^{10} \text{ M}^{-1}$) and its capacity to activate the mast cell at low receptor occupancy may limit the clinical value of these peptides⁸⁶; the design of new competitors with higher affinity, based on more detailed structural information or the random screening of peptide libraries, is a challenging option. An alternative approach is to restrict the accessibility of the binding sites in IgE or FcεRI by stabilizing an inactive conformation of one or the other. It may also be feasible to perturb the interaction of the FcεRI α -chain with its partner subunits, so that IgE binds without inducing signal transduction.

A search has been made for monoclonal antibodies that block binding of IgE to FcεRI without causing mast-cell degranulation^{86,109,110}. The success of such a strategy may depend on the bent structure of IgE, as any epitope involved in FcεRI binding would otherwise have an accessible counterpart on the other side of the molecule. The restricted accessibility of some epitopes may explain why certain anti-Fcε antibodies like soluble FcεRI(α) form only equimolar complexes with free IgE and none when it is bound to its receptor. Moreover, as shown schematically in Fig. 4, antibodies may be able to obstruct the IgE-receptor interaction even when their epitopes do not overlap the receptor binding site. It is even conceivable that by judicious choice of epitopes, a peptide might be found to serve as a vaccine that will generate antibodies that block binding; a candidate peptide has been suggested (Fig. 2; ref. 87). A more conservative approach would be to use univalent Fab or Fv fragments for passive immunization. The possibility of using monoclonal antibodies against receptor peptides in therapy is also being explored³⁸.

Too little is yet known about the signal transduction cascade in mast cells and basophils to allow speculation about how the downstream signalling processes might be interrupted. It is in this area, however, that empirical drug screening has had such notable success. Two fungal products, theophylline and cyclosporin A (CsA), have both been used in the treatment of asthma, the former for several decades, the latter only in recent clinical

trials⁸⁸. Theophylline was originally identified as an antagonist of mast-cell degranulation, but has serious side effects that result from its actions on other cell types. A derivative of theophylline, rolapram, has now been found to act preferentially on mast cells, and has been used to characterize a cyclic AMP phosphodiesterase isozyme, PDE-IV, which is the predominant form in the mast cells of allergic individuals and the target for the drug⁸⁹. Cyclosporin A was originally used to prevent the rejection of allografts. It binds proteins termed cyclophilins and may thus inhibit eosinophil chemotaxis⁹⁰, although other sites of action have been proposed⁹¹; another mode of action is suggested by the recent observation that the drug suppresses expression of CD40-L by mast cells⁸¹. The identification of target proteins for both rolapram and CsA may now allow the development of improved drugs on the basis of structural studies.

Knowledge of the reactions involved in the regulation of IgE synthesis (Fig. 8) may assist in designing drugs to intervene at this level. Monoclonal antibodies against either IL-4 (ref. 92) or the IL-4 receptor⁹³ eliminate expression of IgE in mice, as does inactivation of the IL-4 gene⁹⁴. Similarly, the interactions

between CD40 and CD40-L, and between FcεRII and CR2, can be blocked by soluble ligand fragments or antibodies *in vitro*, and may be exploited therapeutically if a concentration range can be found in which the more general immunosuppressive effects are tolerated^{95,96,111}.

A promising point of purchase is arguably the natural feedback loop by which IgE protects FcεRII against proteolysis. This could be enhanced by the use of a suitable IgE fragment, or by stabilizing the coiled-coil structure of FcεRII directly. Antisense RNA may in the future inhibit transcription of the IgE heavy chain⁹⁷, FcεRII (ref. 98) or FcεRI, but the problems associated with this technique *in vivo* remain to be solved. Chang *et al.* have investigated the use of monoclonal antibodies that crosslink FcγR and mIgE to downregulate IgE synthesis by committed cells⁹⁹. To avoid anaphylaxis caused by crosslinking IgE bound to FcεRI on mast cells, antibodies were raised against the sequence in mIgE that is missing from the secreted form. Antibodies such as these may be the first of a new generation of drugs for the treatment of allergy that are based on knowledge of protein structure. □

- Capron, A., Dessaint, J.-P., Haque, A. & Capron, M. *Prog. Allergy* **31**, 234–267 (1982).
- Capron, A. *et al. Immun. Today* **7**, 15–18 (1986).
- Butterworth, A. E., Dunne, D. W., Fulford, A. J. C. & Thorne, K. J. I. *Immun. Invest.* **21**, 391–407 (1992).
- Capron, A., Dessaint, J. P., Capron, M. & Pierce, R. J. *Immun. Invest.* **21**, 409–422 (1992).
- Wuthrich, B. *Int. Arch. Allergy Appl. Immun.* **90**, 3–10 (1989).
- Barnes, P. J. *Trends biochem. Sci.* **16**, 365–369 (1991).
- Galli, S. J., Gordon, J. R. & Wershil, B. *Curr. Opin. Immun.* **3**, 865–873 (1991).
- Wang, B. *et al. J. exp. Med.* **175**, 1353–1358 (1992).
- Bieber, T. *et al. J. exp. Med.* **175**, 1285–1291 (1992).
- Kikutani, H. & Kishimoto, T. *Res. Immun.* **141**, 249–258 (1990).
- Yokota, A. *et al. Proc. natn. Acad. Sci. U.S.A.* **89**, 5030–5034 (1992).
- Grangette, C. *et al. J. Immun.* **143**, 3580–3588 (1989).
- Bieber, T. *et al. J. exp. Med.* **170**, 309–314 (1989).
- Delespesse, G. *et al. Adv. Immun.* **49**, 149–190 (1991).
- Delespesse, G., Sarfati, M., Wu, C. Y., Fournier, S. & Letellier, M. *Immunol. Rev.* **125**, 78–97 (1992).
- Gordon, J. *Adv. molec. cell. Immun.* (in the press).
- Aubry, J.-P., Pochon, S., Graber, P., Jansen, K. U. & Bonnefoy, J.-Y. *Nature* **358**, 505–507 (1992).
- Bonnefoy, J.-Y., Henchoz, S., Holder, M. J., Hardie, D. & Gordon, J. *Eur. J. Immun.* **23**, 969–972 (1993).
- Bibel, D. J. *Milestones in Immunology* (Science Tech., Madison, 1988).
- Gould, H. J. in *Genetics of Asthma* (eds Holgate S. T. & Church, M. K.) 293–304 (Gower, London, 1992).
- Padian, E. A. & Davies, D. R. *Molec. Immun.* **23**, 1063–1075 (1986).
- Helm, B. A., Ling, Y., Teale, C., Padian, E. A. & Brüggemann, M. *Eur. J. Immun.* **21**, 1543–1548 (1991).
- Beavil, A. J. *et al. Biochem. Soc. Trans.* **21**, 968–972 (1993).
- Nissim, A., Schwarzbaum, S., Siraganian, R. & Eshhar, Z. J. *Immun.* **150**, 1365–1374 (1993).
- Basu, M. *et al. J. biol. Chem.* **268**, 13118–13127 (1993).
- Vercelli, D. *et al. Nature* **338**, 649–650 (1989).
- Truong, M.-J. *et al. J. exp. Med.* **177**, 243–248 (1993).
- Zheng, Y., Shopes, R., Holowka, D. & Baird, B. *Biochemistry* **30**, 9125–9132 (1991).
- Gilfillan, A. M. *et al. J. Immun.* **149**, 2445–2451 (1992).
- Wilson, A. P. M., Pullar, C. E., Camp, A. M. & Helm, B. A. *Eur. J. Immun.* **23**, 240–244 (1993).
- Ra, C. *et al. Internat. Immun.* **5**, 47–54 (1993).
- Ravetch, J. V. & Kinet, J.-P. *A. Rev. Immun.* **9**, 457–492 (1991).
- Metzger, H. *Immunol. Rev.* **125**, 37–48 (1992).
- Kurosaki, T., Gander, L., Wirthmueller, U. & Ravetch, V. J. *exp. Med.* **175**, 447–451 (1992).
- Williams, A. F. & Barclay, A. N. *A. Rev. Immun.* **6**, 381–404 (1988).
- Hogarth, P. M. *et al. Immunol. Rev.* **125**, 21–35 (1992).
- Hulett, M. D., McKenzie, I. F. C. & Hogarth, P. M. *Eur. J. Immun.* **23**, 640–645 (1993).
- Riske, F. *et al. J. biol. Chem.* **266**, 11245–11251 (1991).
- Baniyash, M., Eshhar, Z. & Rivnay, B. J. *Immun.* **136**, 588–593 (1986).
- Ortega, E., Schweitzer-Stenner, R. & Pecht, I. *Biochemistry* **30**, 3473–3483 (1991).
- Springer, T. A. *Nature* **346**, 425–434 (1990).
- Beavil, A. J., Edmeades, R. L., Gould, H. J. & Sutton, B. J. *Proc. nat. Acad. Sci. U.S.A.* **89**, 753–757 (1992).
- Dierks, S. E. *et al. J. Immun.* **150**, 2372–2382 (1993).
- Sarfati, M. *et al. Immunology* **76**, 662–667 (1992).
- Mossalayi, M. D. *et al. EMBO J.* **11**, 4323–4328 (1992).
- Weis, W. I., Kahn, R., Fourme, R., Drickamer, K. & Hendrickson, W. A. *Science* **254**, 1608–1615 (1991).
- Bettler, B., Texido, G., Raggini, S., Rüegg, D. & Hofstetter, H. J. *biol. Chem.* **267**, 185–191 (1992).
- Weis, W. I., Drickamer, K. & Hendrickson, W. A. *Nature* **360**, 127–134 (1992).
- Richards, M. L. & Katz, D. H. J. *Immun.* **144**, 2638–2646 (1990).
- Pochon, S. *et al. J. exp. Med.* **176**, 389–397 (1992).
- Lee, W. T., Rao, M. & Conrad, D. H. J. *Immun.* **139**, 1191–1198 (1987).
- Luo, J., Hofstetter, H., Banchereau, J. & Delespesse, G. J. *Immun.* **146**, 2122–2129 (1991).
- Gould, H. J., Sutton, B. J., Edmeades, R. L. & Beavil, A. J. in *Monographs in Allergy* Vol. 29 (ed. Gordon, J.) 28–49 (Karger, Basel, 1991).
- Gordon, J., Webb, A. J., Guy, G. R. & Walker, L. *Immunology* **60**, 517–521 (1987).
- Pirron, U., Schlunck, T., Prinz, J. C. & Rieber, E. P. *Eur. J. Immun.* **20**, 1547–1551 (1990).
- Speiss, M. *Biochemistry* **29**, 10009–10018 (1990).
- Sugden, B. *Cell* **57**, 5–7 (1989).
- Fearon, D. T. *Curr. Opin. Immun.* **5**, 341–348 (1993).
- Norman, D. G. *et al. J. molec. Biol.* **219**, 717–725 (1991).
- Moore, M. D., DiScipio, R. G., Cooper, N. R. & Nemerow, G. R. *J. biol. Chem.* **264**, 20576–20582 (1989).
- Lowell, C. A. *et al. J. exp. Med.* **170**, 1931–1111946 (1989).
- Carel, J.-C., Myones, B. L., Frazier, B. & Holers, V. M. J. *biol. Chem.* **265**, 12293–12299 (1990).
- Erdei, A., Melchers, F., Schultz, T. & Dierich, M. *Eur. J. Immun.* **15**, 184–188 (1985).
- Esparza, I. *et al. Eur. J. Immun.* **21**, 2829–2838 (1991).
- Carter, R. H. & Fearon, D. T. *Science* **256**, 105–107 (1992).
- Carter, R. H., Tuvson, D. A., Park, D. J., Rhee, S. G. & Gearon, D. T. *J. Immun.* **147**, 3663–3671 (1991).
- Matsumoto, A. K. *et al. J. exp. Med.* **173**, 55–64 (1991).
- Henderson, S. *et al. Cell* **65**, 1107–1115 (1991).
- Wang, F., Kikutani, H., Tsang, S.-F., Kishimoto, T. & Kieff, E. J. *Virology* **65**, 4101–4106 (1991).
- Liu, Y.-L. *et al. Eur. J. Immun.* **21**, 1905–1910 (1991).
- Frade, R., Gauffre, A., Hermann, J. & Barel, M. J. *Immun.* **149**, 3232–3238 (1992).
- Cambier, J. C. & Campbell, K. S. *FASEB J.* **6**, 3207–3217 (1992).
- Clark, M. R. *et al. Science* **258**, 123–126 (1992).
- Eiseman, E. & Bolen, J. B. *Nature* **355**, 78–80 (1992).
- Sugie, K. *et al. Proc. natn. Acad. Sci. U.S.A.* **88**, 9132–9135 (1991).
- Hopkins, C. R. *Trends Biochem. Sci.* **17**, 27–32 (1992).
- Cantley, L. C. *et al. Cell* **64**, 281–302 (1991).
- Vercelli, D. & Geha, R. S. J. *Allergy clin. Immun.* **3**, 285–297 (1991).
- Liu, Y.-J. *et al. Eur. J. Immun.* **21**, 1107–1114 (1991).
- Noelle, R. J., Ledbetter, J. A. & Ruffo, A. *Immun. Today* **13**, 431–433 (1992).
- Gauchat, J.-F. *et al. Nature* **365**, 340–343 (1993).
- Scherer, E., Macy, E., Kimata, H., Gilly, M. & Saxon, A. J. *Immun.* **142**, 481–489 (1989).
- Chen, S.-S. J. *Immun.* **147**, 1581–1586 (1991).
- Kinet, J.-P. *et al. Int. Arch. Allergy appl. Immun.* **94**, 51–55 (1991).
- Blank, U., Ra, C. & Kinet, J.-P. *J. biol. Chem.* **266**, 2639–2646 (1991).
- Kitani, S., Kraft, D., Fischler, C., Mergenhagen, S. E. & Siraganian, R. P. J. *Immun.* **140**, 2585–2588 (1988).
- Stanworth, D. R., Jones, V. M., Lewin, I. V. & Nayyar, S. *Lancet* **336**, 1279–1281 (1990).
- Kay, B. *New Sci.* **137** (1860) 38–42 (1993).
- Chan, S. C. & Hanifin, J. M. J. *Lab. clin. Med.* **121**, 44–51 (1993).
- Erlanger, B. F. *Immun. Today* **14**, 369 (1993).
- Erlanger, B. F. *Immun. Today* **14**, 487–490 (1992).
- Finkelman, F. D. *et al. Proc. nat. Acad. Sci. U.S.A.* **83**, 9675–9678 (1986).
- Finkelman, F. D. *et al. Internat. Immun.* **3**, 599–607 (1991).
- Kuhn, R., Rajewsky, K. & Muller, W. *Science* **254**, 707–710 (1991).
- Hebelle, T., Ahearn, J. M. & Fearon, D. T. *Science* **254**, 102–104 (1991).
- Fanslow, W. C. *et al. J. biol. Chem.* **267**, 655–660 (1992).
- Hall, T. J. & Brostoff, J. *Immunology* **77**, 462–464 (1992).
- Bhatti, L., Behle, K. & Stevens, R. H. *Cell. Immun.* **144**, 117–130 (1992).
- Davis, F. M., Gossett, L. A. & Chang, T.-W. *Biotechnology* **8**, 53–56 (1990).
- Jones, E. Y., Davis, S. J., Williams, A. F., Harlos, K. & Stuart, D. I. *Nature* **360**, 232–239 (1992).
- Shimizu, A. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 1907–1911 (1988).
- Abdelilah, S. G. *et al. Nature* (in the press).
- Prausnitz, C. & Küstner, H. *Zentralblatt für Bakteriologie, Infektionskrankheiten und Hygiene* Abt. 1 **86**, 160–169 (1991).
- Ishizaka, K. & Ishizaka, T. J. *Immun.* **99**, 1187–1198 (1967).
- Liu, F.-T. *Immun. Today* **14**, 486–490 (1993).
- Robertson, M. W. J. *biol. Chem.* **268**, 12736–12743 (1993).
- Mallamaci, M. A. *et al. J. biol. Chem.* **268**, 22076–22083 (1993).
- Tuvson, D. A., Carter, R. H., Soltoff, S. P. & Fearon, D. T. *Science* **260**, 986–989 (1993).
- Davis, F. M. *et al. Springer Sem. Immunopath.* **15**, 51–73 (1993).
- Presta, L. G. *et al. J. Immun.* **151**, 2623–2632 (1993).
- Flores-Romo, L. *et al. Science* **261**, 1038–1041 (1993).
- Bacon, K. *et al. Eur. J. Immun.* **23**, 2721–2724 (1993).

ACKNOWLEDGEMENTS. We thank the following for helpful discussions and for information before publication: J.-Y. Bonnefoy, K. Drickamer, Z. Eshhar, D. Fearon, J. Gordon, W. Hendrickson, Y. Jones, J.-P. Kinet, M. Mossalayi, K. Moulder, C. Page, D. Stuart and W. Weis. The molecular modelling and production of the figures are the work of A. J. Beavil, R. L. Beavil and G. J. Keen. We thank the National Asthma Campaign, the MRC and the SERC for financial support.

The baryon content of galaxy clusters: a challenge to cosmological orthodoxy

Simon D. M. White^{*}, Julio F. Navarro[†], August E. Evrard[‡]
& Carlos S. Frenk[†]

^{*} Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK

[†] Department of Physics, University of Durham, Durham DH1 3LE, UK

[‡] Department of Physics, University of Michigan, Ann Arbor, Michigan 48109, USA

Baryonic matter constitutes a larger fraction of the total mass of rich galaxy clusters than is predicted by a combination of cosmic nucleosynthesis considerations (light-element formation during the Big Bang) and standard inflationary cosmology. This cannot be accounted for by gravitational and dissipative effects during cluster formation. Either the density of the Universe is less than that required for closure, or there is an error in the standard interpretation of element abundances.

A dramatic success of the hot Big Bang theory is its ability to explain the relative abundances of the light elements. Current estimates of the 'primordial' abundances of ^1H , ^2H , ^3He , ^4He and ^7Li are all consistent with those expected a few minutes after the Big Bang, provided that the present universe has baryon density in the range $0.01 < \Omega_b h^2 < 0.015$ (ref. 1). Here, the present value of Hubble's constant, H_0 , is taken to be $100 h \text{ km s}^{-1} \text{ Mpc}^{-1}$, Ω_b is given in units of the critical density needed to close the Universe, and the range is a 95% confidence interval. Baryon densities much larger than this overproduce both ^7Li and ^4He , whereas significantly smaller values overproduce the combined abundance of ^2H and ^3He . Thus, in the context of the standard theory, Ω_b is strongly constrained by the observed abundances of the light elements.

Over the last decade the cosmology community has increasingly accepted the plausibility of the inflation model and of its associated prediction that the Universe should have negligible curvature. In the absence of a cosmological constant this requires the mean matter density, Ω_0 , to be very close to the critical value for closure. The discrepancy between $\Omega_0 = 1$ and $\Omega_b h^2 \sim 0.0125$ is, of course, the main motivation for the hypothesis that most of the mass in the universe is in some non-baryonic and invisible form, for example massive neutrinos or axions². For a long time direct dynamical evidence for such large amounts of dark matter was lacking. Recently, however, comparisons of large-scale motions of galaxies with the density field which induces them have consistently led to large estimates for Ω_0 , often consistent with unity^{3–6}.

Rich clusters of galaxies are the largest objects for which total masses can be estimated directly. The need for unseen dark matter was first identified in such systems⁷, and recent analyses have concluded that stars in the individual galaxies contribute 5–10% of the mass needed to generate the observed orbital motions^{2,8–10}. Baryonic matter is also seen in the form of a hot intracluster medium which emits thermal bremsstrahlung and line radiation at X-ray energies. The density and temperature of this gas can be estimated directly from the surface brightness and spectrum of its X-ray emission, and it has been known for 20 years that its total mass is comparable to that of the galaxies. Recent reviews^{10–12} quote mass fractions for the X-ray-emitting gas in the range $0.04\text{--}0.1 h^{-3/2}$. As some fraction of the dark matter in rich clusters may well also be baryonic, these numbers suggest an approximate lower limit to the baryon fraction in clusters of $\sim 0.05 + 0.04 h^{-3/2}$.

Here we explore the apparent contradiction between the low baryon fraction implied by currently popular cosmological

theories and the larger baryon fractions measured in rich galaxy clusters, a discrepancy which has previously been discussed by a number of authors^{13–18}. The wide field of view and high sensitivity of the Rosat satellite is now producing much improved determinations of the gas distribution in clusters. These are reinforcing previous estimates of the amount of hot gas and, most importantly, are allowing the enclosed gas mass to be measured reliably to much larger radii than before. We use the Coma cluster as an example and derive estimates of the galaxy mass, gas mass, and total masses within a fiducial radius $r_A = 1.5 h^{-1} \text{ Mpc}$, the standard Abell radius. We emphasize that there is no reason to believe that the baryon fraction in this system is in any way anomalous. Indeed, far from the centre, all the rich clusters for which we have so far seen Rosat data bear a close resemblance to Coma (for example, ref. 19).

We then investigate whether the baryon fraction in a cluster can greatly exceed the universal value and conclude that gravitational and dissipative effects during cluster formation cannot account for the apparent discrepancy. We are forced to one of a series of unpalatable conclusions. The mean matter density may fall far short of that required for closure. The mean baryon density may greatly exceed that inferred from light-element abundances. Non-gravitational processes may play a critical role in cluster formation. Standard techniques for estimating cluster masses may be grossly in error. We discuss these possibilities and conclude that neither of the last two appears plausible. If this is accepted, either the Universe has low density or the standard interpretation of element abundances is incorrect.

An inventory of the Coma cluster

We now carry out an inventory of the three main mass components in the Coma cluster: stars in galaxies, hot intracluster gas and dark matter. Present data allow determinations of the mass of each of these components out to the Abell radius. The amount of visible material is probably known to better than 30%; the amount of dark matter can only be determined to similar precision if some specific *a priori* assumption is made about the shape of its distribution.

The mass in stars. The extensive survey of Godwin and collaborators^{20–22} remains the best source of photometry for the Coma cluster. After correction for background galaxies and for cluster galaxies too faint to be catalogued, these data give a total blue luminosity of $1.95 \times 10^{12} h^{-2} L_\odot$ (where $L_\odot$ is the blue luminosity of the Sun) within a circle of radius r_A . Using an analytic fit to the galaxy counts^{23,24}, we find the luminosity within a sphere of radius r_A to be $1.8 \times 10^{12} h^{-2} L_\odot$. To convert lumin-

osity, L , to mass M , we take the relation $M/L = 8.0 h (L/L_*)^{0.35} (M_\odot/L_\odot)$ which van der Marel²⁵ derived from a detailed study of 37 bright elliptical galaxies (here $L_* = 10^{10} h^{-2} L_\odot$ and $M_\odot$ is the mass of the Sun) and we average it over a luminosity function similar to that of Coma to get $\langle M/L \rangle = 6.4 h$. This gives a total stellar mass of

$$M_{\text{gal}} = 1.0 \pm 0.2 \times 10^{13} h^{-1} M_\odot \quad (1)$$

where the uncertainty corresponds to a 0.2 mag uncertainty in the photometry, but neglects possible systematic errors in galaxy mass estimates.

The mass in hot gas. The most accurate determination of the mass in X-ray emitting gas in the Coma cluster comes from the Rosat all-sky survey data. Briel *et al.*²⁶ measured the X-ray surface brightness profile of the cluster to radii well beyond r_A . After trying a wide range of possible shapes for the binding mass profile (each requiring a different gas temperature profile to be in equilibrium and to fit the mean cluster spectrum) they concluded that the gas mass within $2.5 h^{-1}$ Mpc lies in the range $M_{\text{gas}} = 9.0 \pm 2.6 \times 10^{13} h^{-5/2} M_\odot$. At the Abell radius, the surface brightness is still well determined, with an uncertainty of just $\pm 15\%$. Scaling the gas mass of Briel *et al.* to this radius using their analytic fit to the surface brightness profile, we find

$$M_{\text{gas}} = 5.45 \pm 0.98 \times 10^{13} h^{-5/2} M_\odot \quad (2)$$

where we adopt a 15% uncertainty in the Rosat flux and a 10% uncertainty in the conversion from flux to gas mass due to the unknown temperature at large radii. Note that ± 2 standard deviations (s.d.) agrees with the total allowed percentage variation quoted by Briel *et al.*

The dark mass. The total mass within the Abell radius may be estimated independently from optical and from X-ray data. All estimates begin by assuming that the cluster is spherical and in dynamical equilibrium. Optical data constrain the mass strongly only if some assumption is made about the relative distributions of galaxies and of dark matter; for example that the galaxies may be treated as randomly chosen mass elements (the 'light traces mass' assumption). Similarly, in the absence of temperature information at large radii, the X-ray data constrain the mass strongly only if some shape is assumed for the density profile (or equivalently for the temperature profile). Dynamical simulations of cluster formation can eliminate some of these uncertainties; when studying any specific picture for the origin of structure it is reasonable to assume that the density profiles of real clusters are similar in shape to those of simulated clusters.

Comprehensive virial analyses of the optical data on Coma^{24,27} show that if the form of the mass distribution is unconstrained, then the velocity dispersion and number density profiles are consistent with total masses (M_{tot}) within r_A ranging a factor of 2 on either side of the central value, $6.6 \times 10^{14} h^{-1} M_\odot$. But if it is assumed that light traces mass, this quantity is determined to better than 30%:

$$M_{\text{tot}} = 6.7 \pm 1.0 \times 10^{14} h^{-1} M_\odot \quad (3)$$

where we take The and White's total allowed range²⁴ to correspond to ± 2 s.d.

The mass cannot be estimated from the X-ray data directly because there is no reliable temperature information beyond $r_A/3$. In the most thorough analysis to date, Hughes²⁸ finds weak evidence that the gas temperature falls outside the central region (see also ref. 29) and he prefers a model in which the optical light traces the mass. For such a model the X-ray data give a very precise determination of the mass, with a total allowed range of $\pm 13\%$ according to Hughes. Using an analytic fit to the galaxy counts²⁴ to convert Hughes's quoted masses to a value within r_A , we find

$$M_{\text{tot}} = 6.82 \pm 0.34 \times 10^{14} h^{-1} M_\odot \quad (4)$$

where we have taken Hughes's 99% confidence interval to correspond to ± 2.6 s.d. Note the excellent agreement between equations (3) and (4).

The cluster mass profiles found in N -body simulations do not, in general, have the shape implied by the light traces mass assumption. If we assume that real cluster mass profiles have the shape expected in a specific cosmogony, different mass estimates usually result. For simulated clusters we can ask: for what mass would a galaxy population with the observed spatial distribution have the observed velocity dispersion? Or a gas atmosphere with the observed emission profile have the observed mean temperature? We use three sets of simulations to address this problem. All consider cold dark matter (CDM) universes with $\Omega_0 = 1$. The simulations of Frenk *et al.*³⁰ follow the dark matter only, whereas those of J.F.N., C.S.F. and S.D.M.W. (manuscript in preparation) and A.E.E. (manuscript in preparation) also include an adiabatic gas using the smoothed particle hydrodynamics (SPH) technique. In comparing with Coma we adopt a velocity dispersion of 970 km s^{-1} for a magnitude-limited sample of galaxies within the Abell radius^{23,24} and a mean gas temperature of 7.5 keV.

The Frenk *et al.* simulations produced 152 clusters with velocity dispersions greater than 800 km s^{-1} . Near the Abell radius, their mass density profiles have a slope of ~ 2.4 , which is somewhat shallower than the slope (2.72), found for the galaxies in Coma. Thus, we reduce the measured velocity dispersions by $\sqrt{2.72/2.4} = 1.06$ to get estimated 'galaxy' velocity dispersions, $\langle v_g^2 \rangle$, for the simulations. Viewing each artificial cluster along one of three orthogonal directions we can evaluate the ratio of $r_A \langle v_g^2 \rangle / G$ to the mass within r_A (G is the gravitational constant). Using the mean and dispersion in this factor we estimate that in Coma the total mass is given by

$$M_{\text{tot}} = 8.5 \pm 2.6 \times 10^{14} h^{-1} M_\odot \quad (5)$$

The scatter here is quite large and is due to departures from virial equilibrium, to asphericity, and to substructure. For this exercise cluster members were identified using full three-dimensional information. If instead they had been identified in projection, mimicking observational procedures, the inferred uncertainty would have been even larger³⁰.

Our gas dynamical simulations follow 12 clusters with a wide range of final temperatures. In all cases the temperature is nearly constant over the regions of the final cluster that emit most of the X-ray flux, but drops at larger radii. This agrees with the available data in Coma^{28,29}. Also, the surface brightness profiles are very similar in shape to that observed. Scaling these models to a mean X-ray temperature of 7.5 keV gives a mass estimate,

$$M_{\text{tot}} = 1.10 \pm 0.18 \times 10^{15} h^{-1} M_\odot \quad (6)$$

This is consistent with our estimate from the galaxy velocity dispersion for the same assumed profile shape, but its uncertainty is much smaller. This is because X-ray data are less susceptible to projection and contamination effects than optical data, as well as being free from uncertainties due to the unknown shape of galaxy orbits.

We have shown that estimates of the mass within the Abell radius of the Coma cluster from optical and X-ray data agree well for consistent assumptions about the shape of the mass density profile. For profiles similar to those expected in a flat CDM universe, estimated masses are about 50% larger than if light traces mass. To be conservative, and because this is indeed the kind of model we wish to test, we will adopt the CDM estimate, $M_{\text{tot}} = 1.1 \times 10^{15} h^{-1} M_\odot$. We assume a 1 s.d. uncertainty of $\pm 20\%$, slightly larger than the scatter estimated from our relatively small set of simulations.

The baryon fraction. Collecting these results, we find that in the Coma cluster the baryon fraction within the Abell radius is

$$\frac{M_b}{M_{\text{tot}}} \geq 0.009 + 0.050 h^{-3/2} \quad (7)$$

where the first term comes from the galaxies and the second from the gas. The inequality reflects the fact that some of the dark matter may be baryonic. The uncertainty in the right-hand

side is about 25% for our adopted uncertainties in the observational mass estimates. This ratio must be compared with the value, $\Omega_b = 0.0125 h^{-2}$, predicted by cosmic nucleosynthesis. The difference is a factor of 4.7 for $h=1$, and a factor of 3.0 for $h=0.5$. Notice also that the gas mass exceeds the galactic mass by a factor of 5.6 for $h=1$, and by a factor of 16 for $h=0.5$. Although we have focused on the Coma cluster, the available data for other rich clusters suggest that Coma is quite typical (for example, ref. 19).

Maximal baryon infall

We now estimate the maximum baryon enhancement that can be produced by dissipative effects during cluster formation. This upper limit will be a decreasing function of distance from the cluster centre; within a sufficiently large radius the mean baryon fraction must take the global value. Existing simulations suggest that hydrodynamical effects have little influence on, and may even decrease, the baryon fraction within the Abell radius^{31, 33}. To provide a conservative limit we consider (unrealistic) models in which pressure is never able to support the gas against collapse. We first analyse spherical models, and then check their applicability by direct simulation of cluster formation in a CDM universe.

Self-similar accretion. A simple model for cluster formation follows a spherical perturbation in an otherwise uniform universe. Each shell of matter expands to a maximum radius, turns around, falls back, and then settles to equilibrium in the main body of the cluster. For accretion onto a point mass embedded in an Einstein–de Sitter universe, the assumption that the mean equilibrium radius of each shell is a fixed fraction of its maximum radius leads to a cluster density profile^{34, 35}, $\rho \propto r^{-9/4}$. Bertschinger³⁶ derived the full solution to this problem for both collisionless and collisional fluids. In any such model there is a critical radius at which infalling matter either shocks or first meets collisionless material that has passed through the cluster centre. The mean baryon fraction within this, or any larger radius, must take the global value.

As a limiting case, we may treat the baryons as a pressureless fluid that collapses in free-fall onto a central singularity, and then sticks to it. In this case, the similarity solution in an Einstein–de Sitter universe consists of a point mass growing with time t as $t^{2/3}$ surrounded by infalling gas with density profile $\rho \propto r^{-3/2}$ at small radii^{35–37}. For a universe dominated by dark matter but containing a small fraction of such infinitely dissipative ‘baryons’, we can calculate the baryon enhancement within any sphere by assuming the dark matter to follow the collisionless infall solution and the baryons to follow the singular or ‘black hole’ solution. This procedure is easily seen to be conservative for our purposes because it takes the largest possible baryon fraction within any radius, r . In the similarity solutions the radial variable is usually taken to be r/r_{ta} where $r_{\text{ta}}(t)$ is the radius of the shell which is currently turning around. But for our purposes it is more convenient to use the mean enclosed overdensity, $\Delta(r) = 2GM_{\text{tot}}(r)/H_0^2 r^3$, as we can estimate this immediately from the data presented in the last section whereas the determination of r_{ta} for Coma is quite problematic. We use the similarity solutions to express the baryon (M_b) and dark matter (M_{dm}) masses within radius r as a function of the dark matter overdensity at r/r_{ta} , and we estimate the baryon enhancement factor as $Y(\Delta) = M_b(\Delta)/\Omega_b M_{\text{dm}}(\Delta)$.

The thick dashed line in Fig. 1 shows the enhancement calculated in this way from Bertschinger’s solutions³⁶. Our total mass estimate for Coma implies a density contrast of 280 ± 56 within the Abell radius. The maximum possible enhancement at that density is $Y=1.4$, a modest factor compared to that required. Factors $Y>3$ are reached only at densities, $\Delta > 10^4$, corresponding to radii $r < r_A/5$.

3-D gas dynamic simulations. It is reasonable to be suspicious of the spherical symmetry of the similarity solutions. Hierarchical clustering models that fit the overall galaxy distribution form

clusters primarily through mergers within extended filamentary structures, a process which is by no means spherically symmetric. Could this produce greater baryon enhancement than we have just estimated? We address this question by carrying out realistic simulations of cluster formation in a CDM universe, making the extreme assumption that the gas is effectively pressure-free at all times; this is expected to maximize the concentration of baryons into clusters. Previous simulation work has attempted a more realistic treatment of gas physics and has found that while cooling can lead to large baryon condensations in cluster cores, the baryon fraction within the Abell radius remains close to the global value^{18, 31, 33, 38, 39}.

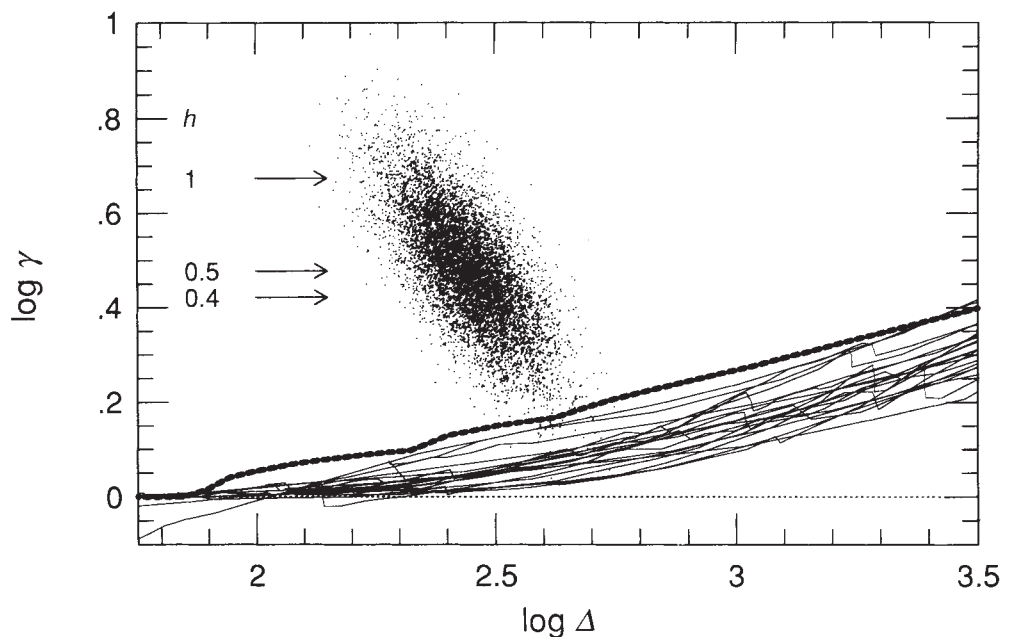
Our simulations use a computer code that combines an N -body integrator with the SPH technique, and uses a nearest-neighbour binary tree to calculate the gravitational forces; it can thus follow the evolution of a mixture of gas and dark matter. (A full description can be found in Navarro and White⁴⁰.) Initial conditions were obtained from the N -body simulations of Frenk *et al.*³⁰, who modelled the growth of structure in a (360 Mpc)³ region of an $\Omega_0=1$, $h=0.5$ CDM universe. From an output time corresponding to fluctuation amplitude $\sigma_8=0.63$ (in the notation of ref. 30), we selected 20 clusters with one-dimensional velocity dispersions $\sigma \sim 1,000 \text{ km s}^{-1}$. For each cluster, the particles in a sphere with overdensity 200 were traced back to the initial conditions and a cubic region containing all of them was defined. This region was then filled with 10,648 new particles on a cubic grid, which were perturbed using both the original waves of the Frenk *et al.* initial conditions and additional waves between their limit and the Nyquist frequency of the new grid. The tidal field due to mass outside this box was provided by $\sim 6,000$ particles of radially increasing mass, using a technique similar to that described by Katz and White³⁸. The gas component is represented by 10,648 particles with the same initial positions and velocities as the dark matter. The gas fraction by mass is 10%, and gas outside the ‘high-resolution’ box is ignored.

To approximate a pressure-free collisional fluid, we adopt an isothermal equation of state with temperature $T=10^4 \text{ K}$, much smaller than the virial temperature of any resolvable structure in the simulations. This results in gas sinking to the centre of any nonlinear structure in one local free-fall time, just as in the self-similar black hole solution. The main difference in the three-dimensional case is that at early times gas concentrates at all the separate local potential minima in the clumpy protocluster. The simulations follow the aggregation of these baryon clumps within the final cluster. We find that most of them merge into a single central clump during, or very soon after, their first pericentric passage.

In Fig. 1 we compare our final clusters with the spherical model by plotting baryon enhancement against mean overdensity, calculated within spheres. The striking result is that all the thin curves representing the simulations lie below the thick curve calculated above. The spherical-infall model solution thus provides an upper limit to the baryon enhancement possible in the inner regions of a cluster. The mergers and the strong deviations from spherical symmetry which occur during cluster formation in a CDM model cannot produce the baryon excess that we observe within the Abell radius of Coma.

Application to the Coma cluster. The ratio, Y , of the baryon fraction in the Coma cluster to the standard nucleosynthesis value for an $\Omega_0=1$ universe is plotted in Fig. 1 at $\Delta=280$. We have assumed $h=0.5$ as any larger value would increase Y still further; arrows indicate values for $h=0.4$ and $h=1$. The ‘observational’ point lies well above the enhancements allowed by our maximally dissipative models. We can quantify the discrepancy by means of Monte Carlo simulations. We generate random values of Y and Δ consistent with the uncertainties in the four ‘observational’ quantities, M_{gal} , M_{gas} , M_{tot} and Ω_b , assuming normal distributions for Ω_b and for the logarithms of the three masses. We take standard deviations equal to the uncertainties quoted in the earlier section on the Coma cluster. We compare

FIG. 1 The baryon overabundance (Y) as a function of mean overdensity (Δ); note the logarithmic scales. The thick dashed curve shows the result of spherical self-similar accretion of pressureless gas in a universe dominated by collisionless dark matter. The thin curves are the results of our N -body-SPH simulations of pressureless accretion in clusters formed in a cold dark matter universe. The dotted line indicates models where no segregation between gas and dark matter has occurred. The dots represent 7,500 (Y , Δ) pairs for the Coma cluster drawn at random using the uncertainties for Ω_b and for the masses of the different components quoted in the text. A Hubble constant of $50 \text{ km s}^{-1} \text{ Mpc}^{-1}$ has been adopted. The arrows show how the value of Y for Coma would change for different values of the Hubble constant, given in units of $100 \text{ km s}^{-1} \text{ Mpc}^{-1}$.



each (Y , Δ) pair with the enhancement curve of a randomly chosen cluster from our 20 simulations. (We plot 7,500 of these pairs as dots in Fig. 1, giving rise to the cloud of points surrounding our best estimate of Y and Δ .) The fraction of such pairs that lies below the curves gives an estimate of the probability that the Coma measurement could be a statistical fluke arising in one of our maximally dissipative CDM models.

Of 10^5 Monte Carlo realizations, only 107 of the (Y , Δ) pairs lie below the model curves. Our models are thus ruled out with about 99.9% confidence. Only five of the pairs give an enhancement less than unity, the typical value found in simulations using 'realistic' parameters for cooling and other dissipative processes. Such models are thus excluded much more strongly than the extreme, maximally segregated case we have considered. Our quoted confidence levels should obviously be treated with scepticism because the uncertainties on the 'observational' quantities are attempts to quantify possible systematic effects rather than true statistical errors. Nevertheless, even if we double the uncertainties quoted above, our maximally segregated model is still excluded with 95% confidence, and the unsegregated model with 98% confidence.

Discussion

The observational data indicate that a large fraction of the binding mass is baryonic in the Coma cluster. For standard values of the Hubble constant ($h=0.5$ – 1) this fraction exceeds the value predicted by cosmic nucleosynthesis in an $\Omega_0=1$ universe by a factor of at least three. Our principal result is that cooling and other dissipative effects cannot enhance the baryon density sufficiently to eliminate this discrepancy if clusters form by gravitational collapse. Although substantial enhancements are possible in the core of a cluster¹⁸, our fiducial radius is only slightly smaller than the boundary that separates infalling matter from matter that is already part of the cluster. Averaged within this boundary, the baryon fraction must take the global value regardless of processes occurring at smaller radii.

What other resolutions of this discrepancy are possible? Let us now consider various possibilities.

Low-density universe. The obvious solution is that $\Omega_0 < 1$. A 'light traces mass' estimate for the total cluster mass seems *a priori* most appropriate in such a universe. The baryon fraction within the Abell radius is then $0.015 + 0.080 h^{-3/2}$. If enhancement effects are weak we may identify this with Ω_b/Ω_0 and

obtain

$$\Omega_0 = 0.16 h^{-1/2} / (1 + 0.19 h^{3/2}) \quad (8)$$

when the nucleosynthesis value is used for Ω_b . This may be too large as some of the unseen matter may be baryonic. The flat universe required by the inflation model can be rescued by a non-zero cosmological constant, a possibility which has other attractive features⁴¹ but which still conflicts with dynamical evidence for large Ω_0 ^{3–6}.

Nonstandard nucleosynthesis. Models for inhomogeneous nucleosynthesis allow some relaxation of the bounds of Walker *et al.*¹. Such models require very carefully tuned parameters in the physics of the quark-hadron phase transition, and the kind of phase transition required currently seems rather unlikely.⁴²

Nongravitational effects. Another possibility is that non-gravitational processes concentrate baryons in clusters. Unfortunately, whereas it is relatively easy to think of reasons why baryons might be under-represented (for example, energy injection by galaxies might lead to gas being blown out) the reverse seems much less likely. If structure on large scales were produced by explosive processes^{43,44}, clusters might form first in baryons and only later accrete dark matter. However, limits on the Compton distortion of the microwave background seem to rule out such models conclusively⁴⁵. One might also imagine a model in which some fraction of the dark matter is kept out of rich clusters by its large random motions. For such models to be consistent with other data, the fraction of the dark matter which is 'hot' cannot exceed 30% (ref. 46), so the discrepancy in baryon fraction could be reduced only by a similar factor.

Mass uncertainties. In our earlier discussion of the Coma cluster, we identified several sources of systematic uncertainty in the masses, and we attempted to quantify their magnitude. In particular, we used numerical simulations to estimate how lack of symmetry, nonequilibrium and projection effects, and the unknown shape of the mass density profile can affect our derived total masses. These effects are serious, particularly when using optical data, but they seem too small to explain the discrepancy. A further complication is the possibility that the gas may be locally inhomogeneous. Let us define a clumping factor, $f = \langle \rho_g^2 \rangle / \langle \rho_g \rangle^2$, where ρ_g is the gas density and the averages are taken over all the gas in some local volume. For fixed X-ray luminosity, X-ray temperature, and for uniform pressure within this volume, the mean gas density and the pressure p scale

approximately as $\langle \rho_g \rangle \propto f^{-1/2}$ and $p \propto f^{1/2}$ respectively. Thus, for given X-ray data, clumping reduces estimates of the gas mass by roughly $f^{-1/2}$. It also affects estimates of the total mass. If the dense regions that dominate the observed emission are strongly coupled to the rest of the gas (for example by magnetic fields), then from the relation $M(r) = -r^2/G\langle \rho_g \rangle dp/dr$, it is clear that the inferred binding mass scales roughly as f . If the dense regions move freely through the hotter gas, the X-ray data do not constrain the binding mass at all. In either case the agreement between binding mass estimates based on optical and on X-

ray data becomes purely fortuitous. In addition, the Sunyaev-Zeldovich decrement⁴⁷ predicted from the X-ray data scales as $f^{1/2} h^{-1/2}$ and agrees with the observed value in Coma for $f \sim 1$ and $h \sim 0.5$ (T. Herbig, personal communication). Thus the available data support the theoretical prejudice that substantial small-scale clumping is implausible.

In conclusion, the large baryon fraction observed in clusters of galaxies may require us to abandon at least one of the basic tenets of current theories for the formation of structure in the universe. □

Received 26 May; accepted 21 October 1993.

1. Walker, T. P. et al. *Astrophys. J.* **376**, 51–69 (1991).
2. Kolb, E. W. & Turner, M. S. *The Early Universe* (Addison Wesley, New York, 1990).
3. Rowan-Robinson, M. et al. *Mon. Not. R. astr. Soc.* **247**, 1–18 (1990).
4. Kaiser, N. et al. *Mon. Not. R. astr. Soc.* **252**, 1–12 (1991).
5. Strauss, M. A. et al. *Astrophys. J.* **397**, 395–419 (1992).
6. Nusser, A. & Dekel, A. *Astrophys. J.* **405**, 437–448 (1993).
7. Zwicky, F. *Helv. phys. Acta* **6**, 110–118 (1933).
8. Dressler, A. *Ann. Rev. Astr. Astrophys.* **22**, 185–221 (1984).
9. Trimble, V. *Ann. Rev. Astr. Astrophys.* **25**, 425–472 (1987).
10. Sarazin, C. L. *X-ray Emissions from Clusters of Galaxies* (Cambridge Univ. Press, 1988).
11. Fabian, A. C., Canizares, C. R. & Nulsen, P. E. J. *Astr. Astrophys. Rev.* **2**, 191–226 (1991).
12. Jones, C. & Forman, W. in *Clusters and Superclusters of Galaxies* (ed. Fabian, A. C.) (Kluwer, Dordrecht, 1992).
13. Shanks, T. *Vistas in Astr.* **28**, 595–609 (1985).
14. White, S. D. M. & Frenk, C. S. *Astrophys. J.* **379**, 52–79 (1991).
15. Fabian, A. C. *Mon. Not. R. astr. Soc.* **253**, 29P–30P (1991).
16. White, S. D. M. in *Clusters and Superclusters of Galaxies* (ed. Fabian, A. C.) 17–28 (Kluwer, Dordrecht, 1992).
17. Makino, N. & Suto, Y. *Publs astr. Soc. Japan* **45**, 13–19 (1993).
18. Babul, A. & Katz, N. *Astrophys. J.* **406**, L51–L54 (1993).
19. Henry, J. P., Briel, U. G. & Nulsen, P. E. J. *Astr. Astrophys.* **271**, 413–424 (1993).
20. Godwin, J. C. & Peach, J. V. *Mon. Not. R. astr. Soc.* **181**, 323–337 (1977).
21. Godwin, J. C., Metcalfe, N. & Peach, J. V. *Mon. Not. R. astr. Soc.* **202**, 113–124 (1983).
22. Millington, S. J. C. & Peach, J. V. *Mon. Not. R. astr. Soc.* **221**, 15–24 (1986).
23. Kent, S. & Gunn, J. *Astr. J.* **87**, 945–971 (1982).
24. The, L. & White, S. D. M. *Astr. J.* **92**, 1248–1253 (1986).

25. van der Marel, R. *Mon. Not. R. astr. Soc.* **253**, 710–726 (1991).
26. Briel, U. G., Henry, J. P. & Böhringer, H. *Astr. Astrophys.* **259**, L31–L34 (1992).
27. Merritt, D. *Astrophys. J.* **313**, 121–135 (1987).
28. Hughes, J. P. *Astrophys. J.* **337**, 21–33 (1989).
29. Watt, M. P. et al. *Mon. Not. R. astr. Soc.* **258**, 738–748 (1992).
30. Frenk, C. S. et al. *Astrophys. J.* **351**, 10–21 (1990).
31. Evrard, A. E. *Astrophys. J.* **363**, 349–366 (1990).
32. Thomas, P. A. & Couchman, H. M. P. *Mon. Not. R. astr. Soc.* **257**, 11–31 (1992).
33. Cen, R. & Ostriker, J. P. *Astrophys. J.* **417**, 404–414 (1993).
34. Gunn, J. *Astrophys. J.* **218**, 592–598 (1977).
35. Gott, J. *Astrophys. J.* **201**, 296–310 (1975).
36. Bertschinger, E. *Astrophys. J. Suppl.* **58**, 39–66 (1985).
37. Bondi, H. *Mon. Not. R. astr. Soc.* **112**, 195–201 (1952).
38. Katz, N. & White, S. D. M. *Astrophys. J.* **412**, 455–478 (1993).
39. Evrard, A. E., Summers, F. & Davis, M. *Astrophys. J.* (in the press).
40. Navarro, J. F. & White, S. D. M. *Mon. Not. R. astr. Soc.* **265**, 271–300 (1993).
41. Efstathiou, G., Sutherland, W. J. & Maddox, S. J. *Nature* **348**, 705 (1990).
42. Fukugita, M. & Hogan, C. *Nature* **354**, 17–18 (1991).
43. Ostriker, J. P. & Cowie, L. L. *Astrophys. J.* **243**, L127–L132 (1981).
44. Ikeuchi, S. *Publs astr. Soc. Japan* **33**, 211–218 (1981).
45. Wright, E. L. et al. *Astrophys. J.* (in the press).
46. Pogosyan, D. Y. & Starobinsky, A. A. *Mon. Not. R. astr. Soc.* **265**, 507–512 (1993).
47. Sunyaev, R. A. & Zeldovich, Ya. B. *Comm. Astrophys. Sp. Phys.* **4**, 173–178 (1972).

ACKNOWLEDGEMENTS. The order of the authors was determined using the first letters of words chosen randomly from the complete poems of T. S. Eliot. We thank N. Metcalfe for help in interpreting the optical photometry of Coma. This work was supported by grants from NASA, SERC and NATO. C.S.F. acknowledges a Sir Derman Christopherson Fellowship.

Mode-switching of a voltage-gated cation channel is mediated by a protein kinase A-regulated tyrosine phosphatase

Gisela F. Wilson* & Leonard K. Kaczmarek

Department of Pharmacology, Yale University, 333 Cedar Street, New Haven, Connecticut 06510, USA

Tyrosine kinases and tyrosine phosphatases are abundant in central nervous system tissue, yet the role of these enzymes in the modulation of neuronal excitability is unknown. Patch-clamp studies of an *Aplysia* voltage-gated cation channel now demonstrate that a tyrosine phosphatase endogenous to excised patches determines both the gating mode of the channel and the response of the channel to protein kinase A. Moreover, a switch in gating modes similar to that triggered by the phosphatase occurs at the onset of a prolonged change in the excitability of *Aplysia* bag cell neurons.

SERINE/THREONINE kinases modulate the activity of both voltage-gated and ligand-gated ion channels in a number of vertebrate and invertebrate systems^{1–6}. In contrast, phosphorylation of ion channels by tyrosine kinases has been observed only for the acetylcholine receptor, a ligand-gated channel⁷. This phosphorylation seems to mediate growth-related changes in receptor localization and may be relatively permanent⁸. Thus, the role of tyrosine phosphorylation in the modulation of the activity of

voltage-gated channels and neuronal signalling is unclear.

In bag cell neurons of the sea-hare, *Aplysia*, elevations in intracellular cyclic AMP trigger a depolarization that causes a roughly 30 min period of spontaneous repetitive action potentials referred to as the afterdischarge^{9–12}. To characterize the mechanism underlying the onset of this spontaneous firing we studied the regulation of a voltage-gated cation channel. We now report that the effect of cAMP-dependent protein kinase (PKA) depends on the gating mode of the channel and that this mode is determined by the phosphorylation state of tyrosine residues on the cation channel or a closely associated protein.

* Present address: Laboratory of Genetics, University of Wisconsin, Madison, Wisconsin 53706, USA.

Application of the T-cell protein tyrosine phosphatase (PTPase)¹³ or activation of a PKA-regulated PTPase endogenous to excised patches causes cation channels to switch from a bursting to a continuously active mode. We also show that when patches containing identified bursters are 'crammed'¹⁴ into the cells of their origin, the PTPase-induced transition in gating modes occurs in response to a stimulus triggering the after-discharge. These studies suggest that tyrosine phosphorylation systems can rapidly modify voltage-gated channels to produce flexible changes in neuronal signalling.

Cation channels show two modes of gating

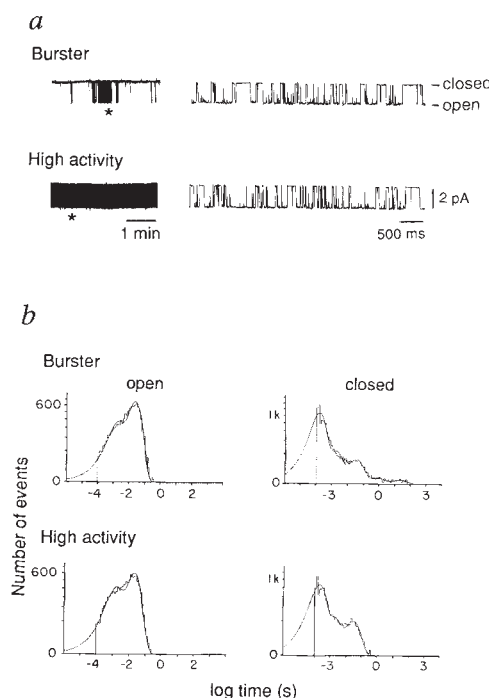
Patch-clamp studies used excised inside-out patches obtained from single bag cell neurons in a resting, unstimulated state. On

patch excision, cation channels were found in either a bursting or high-activity mode. Visual inspection of records representing each of the behavioural phenotypes (Fig. 1*a*, left) suggested that bursting and high-activity channels were distinguished primarily by long-lived closures present in the activity of bursters. Otherwise, the activities of the two phenotypes appeared similar (Fig. 1*a*, right). Because in bursters a single closure could continue in excess of 10 min, the contribution of long closures to burster activity was substantial, despite their infrequency. Indeed, a comparison of the average single channel open probabilities (P_o) in control conditions (Table 1) reveals that the fraction of time bursting channels spent in the open state was about eightfold lower than that for high-activity channels.

Figure 1*b* shows the open- and closed-time histograms, plotted

FIG. 1 Two types of channel activity observed in inside-out patches. *a*, Single-channel records representing the bursting (top) and high-activity modes (bottom). For display, currents were filtered at 500 Hz (−3 dB; 8-pole Bessel). Downward deflections correspond to inward current. Time-expanded traces on the right were taken from the regions indicated by the asterisks. *b*, Duration histograms for the cation channels in *a*. Superimposed curves indicate the exponential fits used to determine time constants. See Table 1 legend.

METHODS. Bag cell neurons from *Aplysia* abdominal ganglia were dissociated and maintained in culture as described previously³⁹. Pipettes contained artificial sea water (in mM): 460 NaCl, 10.4 KCl, 11 CaCl₂, 55 MgCl₂, 10 HEPES, pH 7.8 (NaOH); the bath contained (in mM): 500 K-aspartate, 70 KCl, 0.77 CaCl₂, 1.2 MgCl₂, 10 HEPES, 11 glucose, 0.77 EGTA, 10 glutathione (a reducing agent), 5 ATP, and 0.1 GTP, pH 7.3 (KOH). Under these conditions, the slope conductance, measured between −120 and −20 mV, ranged from 30–36 pS. Currents were recorded using a LIST EPC-7 amplifier, low-pass filtered at 3 kHz and stored on videocassettes. Pipettes were coated with Sylgard and had resistances ranging from 3–10 MΩ. Pipette junction potentials were nulled before seal formation. Cation channels normally were not observed in cell-attached patches, but became visible, most often in clusters, in ~20% of patches after excision. Channels were identified as nonspecific cation channels on the basis of their ion selectivity (Na > Tris > N-methyl-D-glucamine) and voltage-dependence. P_o increased as the holding potential was depolarized from −120 mV and reached maximal levels near −20 mV. The reversal potential extrapolated from currents over this range was between 0 and +10 mV. Because, under certain conditions, currents corresponding to these features can be observed in whole-cell recordings (manuscript in preparation), the absence of cation channel activity from cell-attached patches appears to be a consequence of a local deformation of the membrane by the pipette. Indeed, no rundown was observed in the course of recordings (as might be expected if channels were activated by stretch during excision) and channels continued to be active in



crammed patches (Fig. 4). A similar dependence on structural integrity has been observed for bag cell neuron calcium channels⁴⁰. All experiments were done at a holding potential of −60 mV (corresponding to the resting potential of bag cell neurons) and at ~25 °C.

TABLE 1 Summary of the kinetic features of bursting and high-activity cation channels observed in single-channel patches

Condition	n (N)*	P_o	τ_{o1} (%)	τ_{o2} (%)	τ_{c1} (%)†	τ_{c2} (%)‡	τ_{c3} (%)	τ_{c4} (%)	τ_{c5} (%)
High-activity mode									
Control	8 (76)	0.75 ± 0.06	1.97 ± 0.38 (33)	28.3 ± 4.9 (67)	0.20 ± 0.006 (70)	4.83 ± 1.43 (19)	28.4 ± 4.4 (11)	—	—
PKA	8 (18)	0.45 ± 0.05	1.86 ± 0.33 (38)	19.3 ± 3.1 (62)	0.22 ± 0.003 (65)	6.12 ± 1.88 (18)	46.6 ± 6.1 (16)	—	—
Bursting mode									
Control	12 (53)	0.09 ± 0.03	1.42 ± 0.23 (43)	26.5 ± 6.2 (50)	0.28 ± 0.07 (73)	6.53 ± 2.48 (20)	81.3 ± 18.4 (16)	1770 ± 445 (2)	41925 ± 10365 (0.5)
PKA	3 (8)	0.33 ± 0.17	0.90 ± 0.05 (44)	7.8 ± 1.5 (56)	0.21 ± 0.02 (70)	2.12 ± 0.36 (12)	153.1 ± 93.8 (17)	—	—
Control	1 (53)	0.04	2.87 (38)	33.8 (62)	— (64)	5.78 (10)	67.9 (20)	580 (6)	96200 (0.3)
PTPase	1 (6)	0.59	2.74 (35)	36.6 (65)	— (64)	6.40 (15)	68.0 (21)	—	—
PTPase + PKA	1 (2)	0.37	2.27 (44)	27.2 (56)	— (49)	10.90 (12)	156.0 (39)	—	—

A minimum of 5 and 10 min of recording was used to determine the single channel P_o for high-activity channels and bursting channels, respectively. Currents were sampled at 11 kHz and digitally filtered³⁶ at either 0.5 or 1.5 kHz (gaussian; −3 dB) depending on quality. Durations were determined using a half-amplitude threshold criterion³⁷, corrected for the filter response (equation 17 in ref. 37), and plotted on logarithmic axes to allow better discernment of event populations³⁸. It is likely that the range of burster closures was underestimated, because the most extreme examples of the burster phenotype did not provide a sufficient number of events for analysis. τ values were obtained using the binned maximum likelihood method³⁶. The parameters for the exponentials were varied using a simplex search which was provided with the number of exponentials at the start. All τ values are expressed in ms, and are followed by the percentage of events described by each exponential given in parentheses. Standard errors for the average P_o and τ are given where applicable.

* The n given represents the number of single channel patches used for analysis. The total number of patches (N) includes both single and multiple channel patches. Unless otherwise noted in the text, qualitatively similar results were obtained in all patches in that similar changes in P_o and bursting were observed.

† The briefest closures are poorly resolved at the filter frequency used, leading to an overestimation of τ_{c1} , as well as potentially obscuring any differences between the two phenotypes in gating events on the μ s timescale. In an attempt to minimize this difficulty, values obtained from channels filtered at 0.5 kHz were not included in the determination of the mean τ_{c1} . Therefore, the n for this column are, in descending order: 3, 3, 8, 3.

‡ A somewhat improved fit was obtained if the closed durations currently described only by τ_{c2} (corresponding to the trough between the first and second peaks) were fit with two exponentials instead of one. Variability in the τ_{c2} column may be due to the conservative fit chosen.

on logarithmic scales, for the channels of Fig. 1a. When the open-time distributions (left) were compared, no difference between bursting and high-activity channels was evident. Both open-time distributions were best fit by a sum of two exponentials, τ_{o1} and τ_{o2} , with a higher percentage of events falling under the second, slower exponential (Table 1). If, as is generally assumed, the number of exponentials required to fit these distributions reflects the number of states a channel may enter¹⁵, these results suggest that both phenotypes have two open states.

Comparison of the closed-time histograms for the two channel phenotypes (Fig. 1b, right) revealed that the closed times of burster channels were distributed over a much wider range than those of high-activity channels. High-activity channels never exhibited closed times longer than 1 s, whereas bursters exhibited an additional population of closed times extending over the 1–100 s range. Closed-time distributions of high-activity channels were fitted well by a sum of three exponentials (Table 1, τ_{c1} – τ_{c3}), whereas fits of the closed-time distributions of burster channels required a sum of five exponentials (Table 1, τ_{c1} – τ_{c5}). With the exception that the τ_{c3} values for bursting channels may be larger, the average values for τ_{o1} – τ_{c3} seem to be similar for the two phenotypes. These results suggest that the primary kinetic feature distinguishing the bursting and high-activity modes is the presence or absence of the fourth and fifth closed states.

PKA effect depends on gating mode

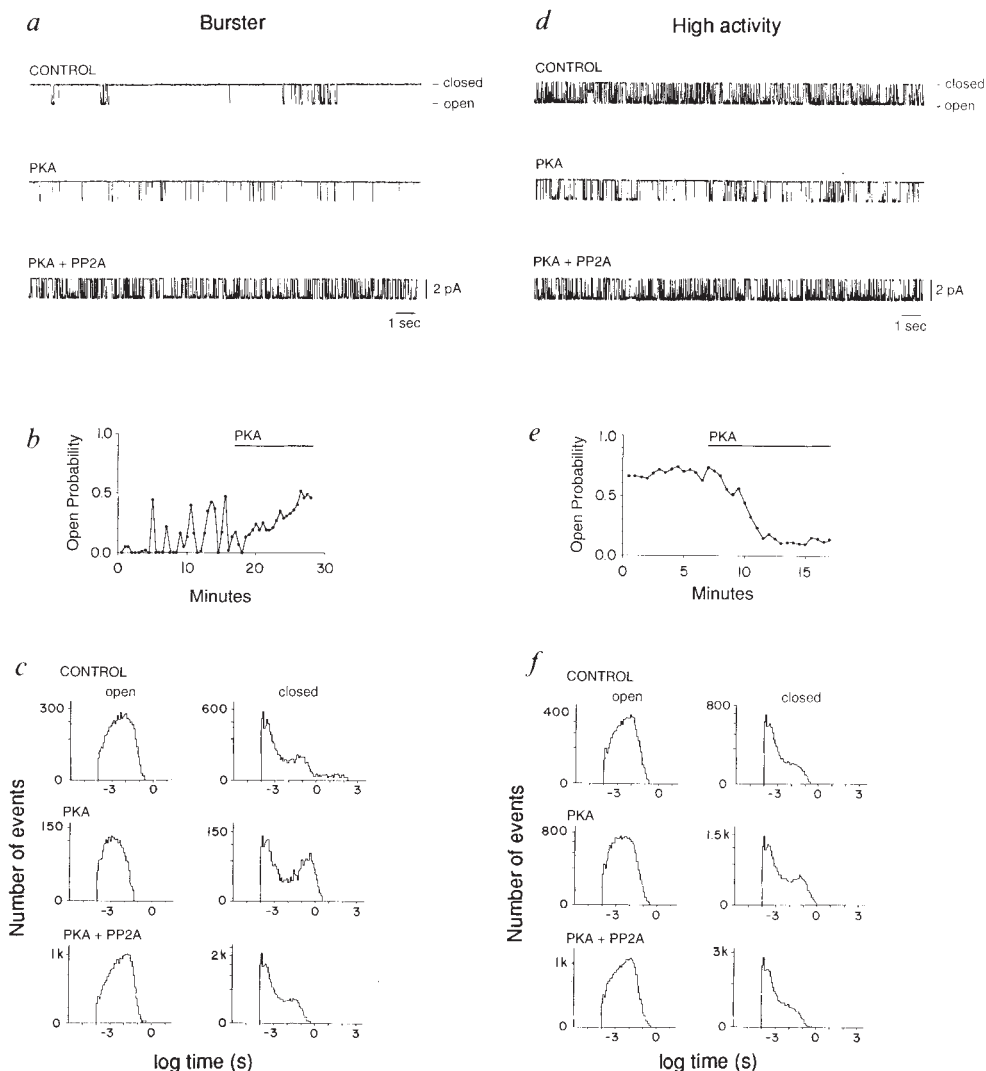
We next examined the effect of the catalytic subunit of PKA on

both cation channel phenotypes. As shown for the single channel patch in Fig. 2a (top and middle traces), the catalytic subunit of PKA (100 nM) applied to bursting channels increased the fraction of time channels spent in the open state (see also Table 1) by removing the long-lived closures. In control conditions, large fluctuations in P_o were observed (Fig. 2b), reflecting the contribution of the long-lived closures to burster activity. PKA eliminated these fluctuations within minutes of its application ($n = 8$ of 10 patches). No change in P_o or bursting was observed when PKA was applied in the presence of the Walsh inhibitor (1 μ M) in 3 of the 4 patches tested (data not shown). In the remaining patch, a long delay (16 min) before the onset of the effect was observed.

Figure 2c shows the duration histograms for the channel in Fig. 2a before and after application of PKA. In the presence of PKA, channel openings were shorter as indicated by the decreases in the relative number of events and the time constant for the second open state (τ_{o2}). PKA also lengthened the duration of most closures as indicated by the increases in τ_{c3} (corresponding to the second peak). The most striking PKA-induced change, however, was the elimination of events described by τ_{c4} and τ_{c5} . Only this latter change accounts for the PKA-induced increase in P_o .

PKA (20–100 nM) applied to high-activity channels produced changes in P_o that were opposite in direction to those observed in bursters ($n = 18$). PKA decreased the fraction of time channels spent in the open state (Fig. 2d, top and middle traces). No

FIG. 2 PKA and PP2A effects on cation channels in bursting and high-activity modes. **a** and **d**, Representative records from bursting (**a**) and high-activity (**d**) channels in control conditions (top), immediately after application of 100 nM of the catalytic subunit of PKA (Sigma, middle), and subsequently, 12 nM PP2A (bottom). PP2A would be expected to fail to restore bursting if the PKA effect was indirect (for example, if PKA activates a phosphatase and the opposing kinase is absent in the excised patch). For display, currents were low-pass filtered at a cut-off frequency of 500 Hz. As shown in **d**, half-amplitude sub-conductance states were observed, but were rare and appeared to occur with equal frequency in high activity and burster channels (see also Fig. 3f, top). **b** and **e**, Time course of the change in P_o observed in response to PKA for representative bursting and high-activity channels, respectively. Data points represent the P_o obtained for successive 30-s intervals. An effect was noticeable within 2 min of PKA application for both channels, although about 5 min was required for the P_o to stabilize at its new value. **c** and **f**, Open- and closed-time distributions for the bursting and high-activity channels shown in **a** and **d** in control conditions (top), after application of PKA (middle) and, subsequently, PP2A (bottom). PP2A was provided by the D. L. Brautigam laboratory.



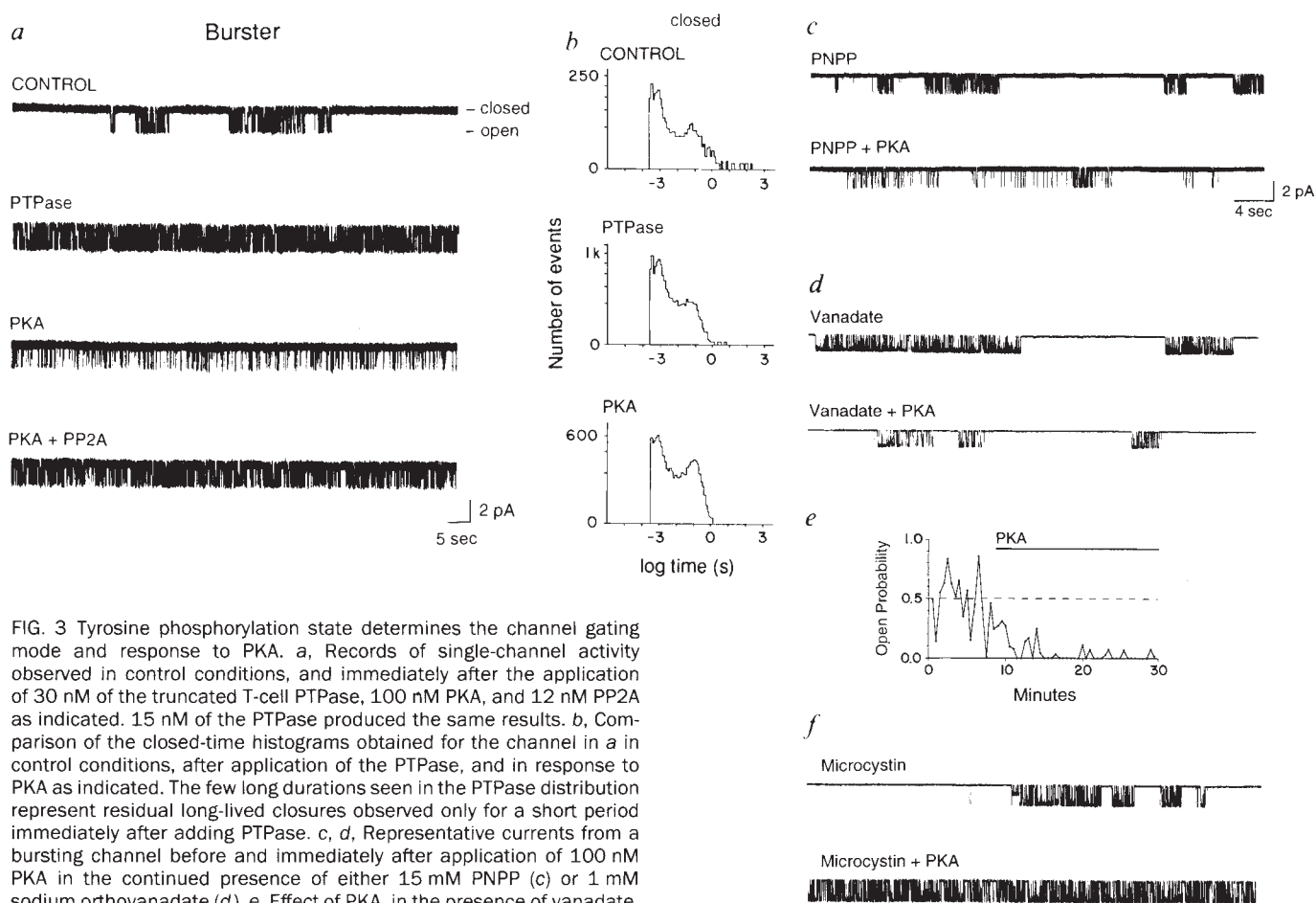


FIG. 3 Tyrosine phosphorylation state determines the channel gating mode and response to PKA. *a*, Records of single-channel activity observed in control conditions, and immediately after the application of 30 nM of the truncated T-cell PTPase, 100 nM PKA, and 12 nM PP2A as indicated. 15 nM of the PTPase produced the same results. *b*, Comparison of the closed-time histograms obtained for the channel in *a* in control conditions, after application of the PTPase, and in response to PKA as indicated. The few long durations seen in the PTPase distribution represent residual long-lived closures observed only for a short period immediately after adding PTPase. *c*, *d*, Representative currents from a bursting channel before and immediately after application of 100 nM PKA in the continued presence of either 15 mM PNPP (*c*) or 1 mM sodium orthovanadate (*d*). *e*, Effect of PKA, in the presence of vanadate, on the single-channel P_o of the bursting channel shown in *d*. Data points represent the P_o for successive 30-s intervals. *f*, Representative currents from a bursting channel before and immediately after application of 100 nM PKA in the continued presence of 2 nM microcystin-LR (Calbiochem). In all cases, currents were low-pass filtered at a cut-off frequency of 500 Hz for display. Although spontaneous transitions from

the bursting to high-activity mode were occasionally observed, in no case did we observe the reverse transition from high-activity to bursting modes. The truncated form of the T-cell PTPase is missing an 11K C-terminal segment which confers specificity¹³ and was a gift from the E. H. Fischer laboratory.

decrease in P_o was observed when PKA was applied in the absence of ATP ($n=3$, data not shown) or in the presence of the Walsh inhibitor (1 μ M, $n=6$, data not shown). The PKA-induced decrease in P_o appeared to be accounted for by decreases in τ_{o2} and increases in τ_{c3} (Fig. 2*f*; Table 1). These results suggest that, apart from PKA-induced removal of the fourth and fifth closed states observed only in bursters, the effects of PKA on high-activity and burster kinetics are similar.

If PKA phosphorylates cation channels directly, then serine/threonine-directed phosphatases should reverse the PKA effect. For high-activity channels ($n=2$), we found that application of protein phosphatase 2A (PP2A; 12 nM)^{16–18} caused channel recordings, as well as duration histograms, to assume a pre-PKA appearance (Fig. 2*d*, *f*, bottom). In contrast, application of PP2A to PKA-treated bursters did not restore bursting (Fig. 2*a*, bottom), but rather increased the P_o ($n=2$). Although closures longer than 1 s remained absent, in the presence of PP2A all other burster features returned to control levels (Fig. 2*c*, bottom) in a manner similar to that observed for high-activity channels. Indeed, in the presence of PP2A, the P_o for the two channels was nearly the same (0.77 and 0.72 for the high-activity and burster channels, respectively).

Tyrosine phosphorylation determines mode

During our experiments, we occasionally observed bursting channels switch spontaneously to a high activity-like mode, so

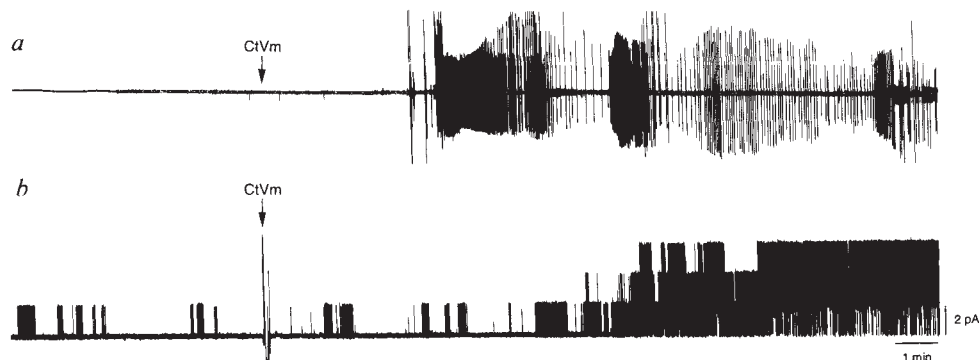
we investigated the possibility that the difference between channel phenotypes was a consequence of channel modification. We found the difference between phenotypes to be due to tyrosine phosphorylation. Application of the T-cell PTPase caused bursters to assume a high activity appearance (Fig. 3*a*, $n=6$). Treatment with the T-cell PTPase eliminated the long-lived closed states previously described by τ_{c4} and τ_{c5} (Fig. 3*b*), whereas no other kinetic features appeared to be affected (Table 1). In contrast, no change in burster activity was observed when the T-cell PTPase was applied in the presence of sodium orthovanadate (1 mM), a PTPase inhibitor ($n=5$).

Application of the T-cell PTPase also changed the response of bursters to subsequently applied PKA ($n=2$). PKA (100 nM) now decreased the P_o (Fig. 3*a*, third trace). As for high-activity channels, the decrease in P_o was accounted for by decreases in τ_{o2} and increases in τ_{c3} (Fig. 3*b*, bottom; Table 1). In addition, application of PP2A now appeared to reverse the effect of PKA (Fig. 3*a*, bottom).

PTPase inhibitors block the effect of PKA

Because PKA and the T-cell PTPase both removed the long-lived closures present in bursting channels, one action of PKA might be to activate a tyrosine phosphatase endogenous to excised patches. Such an action would be consistent with the failure of PP2A to restore bursting. We therefore examined the effect of PKA on bursting channels in the presence of phosphat-

FIG. 4 Effect of CtVm on the spontaneous activity of bag cell neurons and gating mode of bursting channels. *a*, Extracellular recording of bag cell neuron activity in an acutely isolated abdominal ganglion. Abdominal ganglia were isolated and neuronal activity recorded as described previously^{10,12}. Application of CtVm ($100 \mu\text{g ml}^{-1}$) caused normally silent bag cell neuron clusters to discharge spontaneous action potentials for ~ 30 min ($n = 6$). As is the case for other stimuli that elicit an afterdischarge^{9-10,12}, no effect was observed when CtVm was added during the subsequent



~ 18 h refractory period (not shown). *b*, Recording from a crammed patch containing identified bursting channels. Application of CtVm ($100 \mu\text{g ml}^{-1}$) caused the bursters to transit to the high-activity mode. For the patch shown, only one current level had been observed for several minutes in control conditions, however, 3 current levels became evident after application of CtVm. Because it seems unlikely that the 3 channels were active without overlap in the control recording, the

simplest explanation is that CtVm also recruited 2 previously silent channels. Each current level appeared to pass through a stage of bursting before becoming continuously active. The apparent recruitment of additional channels also could occur in excised patch experiments in response to PKA and the T-cell PTPase, but was never observed in the presence of PTPase inhibitors. CtVm was a gift from the B. Olivera laboratory. Time bar applies to *a* and *b*.

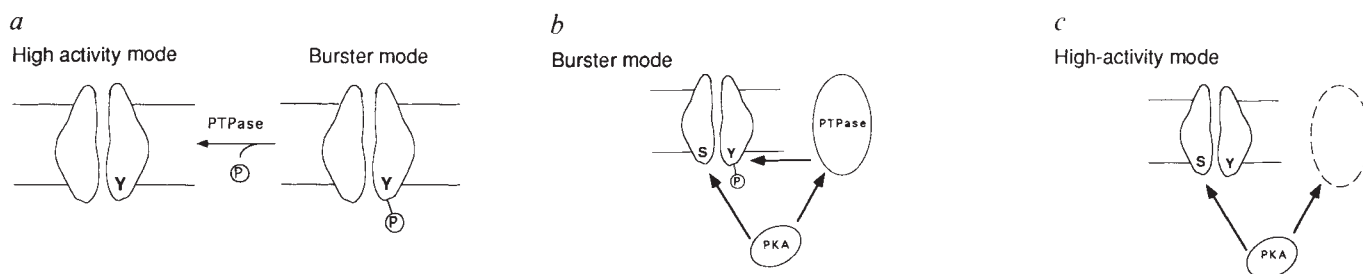


FIG. 5 Model of PTPase and PKA effects on cation channels. *a*, the distinction between bursting and high-activity modes is a consequence of tyrosine phosphorylation. Activated PTPases cause cation channels switch from bursting to high-activity modes by dephosphorylating tyrosine residues on the cation channel or a closely associated protein. *b*, The observed effect of PKA on cation channels in the bursting mode is

mediated by 2 separate phosphorylation reactions: (1) PKA activates the endogenous PTPase, and (2) PKA phosphorylates the cation channel or a closely associated protein. *c*, The observed effect of PKA on cation channels in high-activity mode is mediated solely by the phosphorylation of cation channels (or closely associated proteins). PKA may still activate the PTPase, but there is no phosphate group for the PTPase to remove.

ase inhibitors. PKA applied in the presence of either para-nitrophenyl-phosphate (PNPP, $n = 2$) or sodium orthovanadate ($n = 6$), two tyrosine phosphatase inhibitors, failed to remove the long-lived closures of bursting channels (Fig. 3*c, d*). In fact, in 7 of the 8 patches tested, PKA decreased the P_o (Fig. 3*e*) in a manner similar to that observed for channels in the high-activity mode. In contrast, microcystin (2–10 nM), a specific inhibitor of serine/threonine phosphatases¹⁹, failed to inhibit the removal of bursting (Fig. 3*f*; $n = 6$). These results suggest that PKA removes bursting by activating a PTPase endogenous to excised patches.

Neuronal stimulation alters gating mode

In response to brief stimulation, bag cell neurons undergo a 30-min after-discharge followed by about an 18-hour refractory period⁹. This prolonged change in excitability, which leads to egg-laying behaviours, can be triggered in intact ganglia by a variety of agents^{10-12,20} including venom from the marine snail, *Conus textile* (CtVm; Fig. 4*a*). In isolated neurons, this response is mediated by the rapid activation of a depolarizing current (manuscript in preparation). To determine whether tyrosine dephosphorylation occurs at the onset of the afterdischarge, we excised patches from unstimulated bag cell neurons to identify the channel phenotype. Patches containing bursters then were reinserted ('crammed'¹⁴) into neurons so that the intracellular

channel face was in contact with the cytoplasm. Subsequent bath application of CtVm triggered the transition from bursting to high-activity mode as observed previously for excised patches in response to application of either PKA or the T-cell PTPase (Fig. 4*b*; $n = 5$ of 6 patches). This effect occurred over a time course similar to that required for initiation of the after-discharge and demonstrates that the PTPase-induced transition in gating is an early event in the response of bag cell neurons.

Discussion

This study illustrates that tyrosine kinases and phosphatases can play a key role in the regulation of voltage-gated ion channel activity and, consequently, neuronal output and firing patterns. For cation channels of bag cell neurons, phosphorylation on tyrosines (Fig. 5*a*) promotes bursting by permitting channels to enter long-lived closed states otherwise not evident in channel recordings.

Our data also suggest that the phosphorylation state of tyrosine residues determines whether the observed channel response to PKA will be an increase or decrease in P_o . If the tyrosine site is occupied, as is the case for bursting channels (Fig. 5*b*), then activation of the endogenous PTPase by PKA will produce the dephosphorylation-mediated switch in gating modes. In contrast, in the case of high-activity channels (Fig. 5*c*), activation of the endogenous PTPase by PKA may occur, but will not be

reflected by changes in cation channel activity because there is no phosphate group bound to the tyrosine site. Instead, PKA will produce only changes in τ_{o2} and τ_{e3} which, alone, result in a decreased P_o . Finally, the range of P_o s following PKA treatment (Figs 2a, d and 3f) and the low basal PTPase activity lead us to suspect additional regulation of both cation channel and PTPase. Both ligand binding and interaction with cytoskeletal elements can regulate the activity of other PTPases^{21,22}.

The PTPase endogenous to bag cell neurons is observed in most excised patches and thus, may be a member of the transmembrane class of PTPases, including CD45^{23,24} or of the subset of cytoplasmic PTPases which possess an SH2 domain or other means of association with membrane bound proteins²⁵⁻²⁸. Previous work has demonstrated that dephosphorylation of CD45 serine residues results in a decrease in dephosphorylating activity²⁹, however, the kinase responsible for phosphorylating these residues has not been identified. To our knowledge, this study provides the first example of the regulation of PTPase activity by PKA in a cell-free preparation. Isolation of the bag

cell neuron PTPase will be required to determine whether PKA contacts the PTPase directly.

Cation currents bearing similarities to the channel in this study have been observed in neurons of a number of mammalian systems, as well as in invertebrate neurons (see refs 30-35 for example). Like the cation channel, these currents are largely non-inactivating, are selective for sodium over large monovalent cations, and are thought to underlie neuronal bursting. We speculate that the dephosphorylation effects described here underlie the depolarization and after-discharge induced by cAMP in bag cell neurons. In this scheme, elevations in intracellular cAMP, through activation of PKA and the PTPase, stimulate the removal of phosphate groups from tyrosine residues, thereby increasing cation channel activity and triggering the after-discharge. The opposing effects of PKA on channels in the bursting versus high-activity modes also bear an intriguing resemblance to the different effects of cAMP analogues in previously unstimulated and refractory bag cell neurons¹⁰⁻¹² and may account for the diversity in the effects of cAMP on cation currents of other systems. □

Received 17 June; accepted 14 October 1993.

- Kaczmarek, L. K. & Levitan, I. (eds) *Neuromodulation: The Biochemical Control of Neuronal Excitability* (Oxford Univ. Press, New York, 1987).
- Byrne, J. H., Zwartjes, R., Homayouni, R., Critz, S. D. & Eskin, A. *Adv. Sec. Mess. Phosphop. Res.* **27**, 47-108 (1993).
- Levitan, I. *J. Membr. Biol.* **87**, 177-190 (1985).
- Nicoli, R. *Science* **241**, 545-551 (1988).
- Swope, S. L., Moss, S. J., Blackstone, C. D. & Huganir, R. L. *FASEB J.* **6**, 2514-2523 (1992).
- West, J. W., Numann, R., Murphy, B. J., Scheuer, T. & Catterall, W. A. *Science* **254**, 866-868 (1991).
- Huganir, R. L., Miles, K. & Greengard, P. *Proc. natn. Acad. Sci. U.S.A.* **81**, 6968-6972 (1984).
- Wallace, B. G., Qu, Z. & Huganir, R. L. *Neuron* **6**, 869-878 (1991).
- Kupfermann, I. & Kandel, E. R. *J. Neurophysiol.* **33**, 865-876 (1970).
- Kaczmarek, L. K., Jennings, K. & Strumwasser, F. *Proc. natn. Acad. Sci. U.S.A.* **75**, 5200-5204 (1978).
- Kaczmarek, L. K. & Strumwasser, F. *J. Neurosci.* **1**, 626-634 (1981).
- Kauer, J. A. & Kaczmarek, L. K. *J. Neurosci.* **5**, 1339-1345 (1985).
- Zander, N. F. et al. *Biochemistry* **30**, 6964-6970 (1991).
- Kramer, R. H. *Neuron* **2**, 335-341 (1990).
- Hille, B. *Ionic Channels of Excitable Membranes* (Sinauer, Sunderland, 1992).
- Pallas, D. C. et al. *Cell* **60**, 167-176 (1990).
- Chung, S., Reinhart, P. H., Martin, B. L., Brautigam, D. & Levitan, I. B. *Science* **253**, 560-562 (1991).
- Reinhart, P. H., Chung, S., Martin, B. L., Brautigam, D. L. & Levitan, I. B. *J. Neurosci.* **11**, 1627-1635 (1991).
- Suganuma, M. et al. *Toxicon* **30**, 873-878 (1992).
- Heller, E., Kaczmarek, L. K., Hunkapiller, M. W., Hood, L. E. & Strumwasser, F. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2328-2332 (1980).
- Desai, D. M., Sap, J., Schlessinger, J. & Weiss, A. *Cell* **73**, 541-554 (1993).
- Lokeshwar, V. B. & Bourguignon, L. Y. M. *J. Biol. Chem.* **267**, 21551-21557 (1992).
- Charbonneau, H., Tonks, N. K., Walsh, K. A. & Fischer, E. H. *Proc. natn. Acad. Sci. U.S.A.* **85**, 7182-7186 (1988).
- Alexander, D. R. *New Biol.* **2**, 1049-1062 (1990).
- Shen, S.-H., Bastien, L., Posner, B. I. & Chretien, P. *Nature* **352**, 736-739 (1991).
- Guan, K., Haun, R., Watson, S. J., Geahlen, R. L. & Dixon, J. E. *Proc. natn. Acad. Sci. U.S.A.* **87**, 1501-1505 (1990).
- Gu, M., York, J. D., Warshawsky, I. & Majerus, P. W. *Proc. natn. Acad. Sci. U.S.A.* **88**, 5867-5871 (1991).
- Yang, Q. & Tonks, N. K. *Proc. natn. Acad. Sci. U.S.A.* **88**, 5949-5953 (1991).
- Ostergaard, H. L. & Trowbridge, I. S. *Science* **253**, 1423-1425 (1991).
- Alonso, A. & Llinas, R. R. *Nature* **342**, 175-177 (1989).
- Llinas, R. & Sugimori, M. *J. Physiol., Lond.* **305**, 171-195 (1980).
- Hoehn, K., Watson, T. W. J. & MacVicar, B. A. *Neuron* **10**, 543-552 (1993).
- Strafstrom, C. E., Schwandt, P. C., Chubb, M. C. & Crill, W. E. *J. Neurophysiol.* **53**, 153-170 (1985).
- Green, D. J. & Gillette, R. *Nature* **306**, 784-785 (1983).
- Partridge, L. D. & Swandulla, D. *Pflug. Arch.* **410**, 627-631 (1987).
- Sigworth, F. J. in *Single-channel Recording* (eds Sakmann, B. & Neher, E.) 301-321 (Plenum, New York, 1983).
- Colquhoun, D. & Sigworth, F. J. in *Single-channel Recording* (eds Sakmann, B. & Neher, E.) 191-263 (Plenum, New York, 1983).
- Sigworth, F. J. & Sine, S. M. *Biophys. J.* **52**, 1047-1054 (1987).
- Kaczmarek, L. K., Finbow, M., Revel, J. P. & Strumwasser, F. *J. Neurobiol.* **10**, 535-550 (1979).
- Strong, J. A., Fox, A. P., Tsien, R. W. & Kaczmarek, L. K. *Nature* **325**, 714-717 (1987).

ACKNOWLEDGEMENTS. We thank D. L. Brautigam, E. H. Fischer and B. Olivera for the gifts of PP2A, the T-cell PTPase, and CtVm, respectively, F. J. Sigworth for providing software and assistance in the kinetic analysis and E. M. Blumenthal and S. Chung for discussions and reading the manuscript. This work was supported by an NIH postdoctoral fellowship (to G.W.) and grants from the NIH (to L.K.).

Laboratory evidence for highly unsaturated hydrocarbons as carriers of some of the diffuse interstellar bands

J. Fulara, D. Lessen, P. Freivogel & J. P. Maler

Institut für Physikalische Chemie, Universität Basel,
Klingelbergstrasse 80, CH-4056 Basel, Switzerland

THERE are many absorption lines in the visible and near-infrared spectra of stars located on the far side of diffuse interstellar clouds. The origin of these 'diffuse interstellar bands' (DIBs) has remained an unanswered question since their discovery almost 70 years ago^{1,2}. There are now over 100 known bands^{1,3-6} and it is clear from the range of line widths, depths and shapes that the lines are unlikely to come from a single 'carrier'. Many of the proposed carriers, such as gas-phase carbon chains⁷, fullerenes⁸ and dust grains⁹, fail in having ultraviolet absorption lines where none has yet been observed in the stellar spectra. Polycyclic aromatic species such as $C_{16}H_{10}^+$ (ref. 10) and $C_{10}H_8$ (ref. 11) were recently claimed to be good candidates for carriers of some of the DIBs. Here we present laboratory evidence that highly unsaturated hydrocarbons with carbon numbers 6–12 may be the carriers of some of the DIBs in the range 480–1,000 nm. We deposit mass-selected molecules in a neon matrix at 5 K and measure their near-infrared, visible and ultraviolet spectra. Not only do these species have visible and near-infrared lines corresponding to fifteen DIBs, but they also show no absorption lines in the ultraviolet, consistent with astronomical observations.

Special conditions are required to generate highly unsaturated hydrocarbons for spectroscopic study. The experimental procedure is based on the measurement of the electronic absorption spectra of mass-selected species in neon matrices. Negative ions are initially produced, mass-selected and codeposited with neon, while irradiating with photons, to produce a matrix at 5 K. Predominantly neutral species are produced because the anions lose an electron by photon or collisional detachment during deposition.

Anions of $C_nH_m^-$, where $n=2-12$ and $m<3$, were generated in a discharge-type ion source¹² from a mixture of 10% diacetylene in argon. Species containing odd numbers of carbon atoms were not produced in sufficient abundance for such experiments. Mass resolution was too low to distinguish the number of hydrogen atoms in a selected $C_nH_m^-$ species, however measurements on a higher-resolution instrument (equipped with an electron impact source) show that for $n=4, 6$ the monohydride ions C_nH^- are much more abundant than C_n^- or $C_nH_2^-$.

Mass-selected $C_nH_m^-$ ions (typical current 3–7 nA) were co-deposited with neon (ratio 1:10⁶) on a rhodium-coated copper mirror held at 5K, until a matrix of thickness $\sim 100 \mu m$ was grown. During this process the neon surface was irradiated (~ 50 mW) from a high-pressure Hg lamp, which was found to enhance the intensity of the observed bands. After deposition the absorption spectrum was measured in the 230–1,150 nm region by passing light through the thin side of the matrix along its length (~ 20 mm)¹³. Because of the increasing degree of scattering towards the ultraviolet, the spectra had to be measured in sections.

The observed spectra are displayed in Figs 1 and 2. The shape of the transmission curve is determined by the light source, detector and matrix quality. Four traces (A to D) correspond to mass-selection of $C_nH_m^-$, with $n=6, 8, 10$ or 12 , respectively. In the case of $C_4H_m^-$ selection, no absorption was detected in the 460–1,150 nm range and for $n>12$ the attainable current was too small.

In the ultraviolet range a common feature is the C_3 band

system, with origin at 405.8 nm and extending to 370 nm (Fig. 1). Spectra resulting from the selection of $C_6H_m^-$ and $C_8H_m^-$ show no other absorption down to 230 nm. For $n=10$ and 12 , broad absorption peaks below 320 or 370 nm are seen (Fig. 1). These only appear for $n \geq 10$, and are probably due to cyclic isomers (that is, resembling aromatic systems which absorb in this region). In fact several experimental and theoretical studies on the C_n species show clearly that around $n=10$, cyclic isomers are also formed¹⁴.

In the visible part of the spectrum (400–600 nm) a few distinct, relatively strong absorption bands are apparent, especially when $C_8H_m^-$ is selected (Fig. 2a). The bands which are distinct in the spectra are indicated by the numbered vertical lines at the top of the figure and the wavelength of their maxima are collected in Table 1 (where their appearance in the respective experiments is identified by A, B, C, D). Two stronger and two weaker bands, marked with a star in Fig. 2a lie close to observed DIBs (Table 1). The listed DIBs by Herbig¹ in this region are plotted as sticks at the bottom of Fig. 2a.

The C_3 , C_2^+ , C_2^- bands appearing in the spectra indicate that fragmentation processes take place. (Loss of C_3 suggests that some of the weaker bands may be due to species of the type C_nH_m , where n is odd). Fragmentation is driven in part by the kinetic energy of the ions (~ 150 eV) and by radiation from the Hg lamp. The presence of many weaker bands in the spectra indicates that several species are generated by these processes.

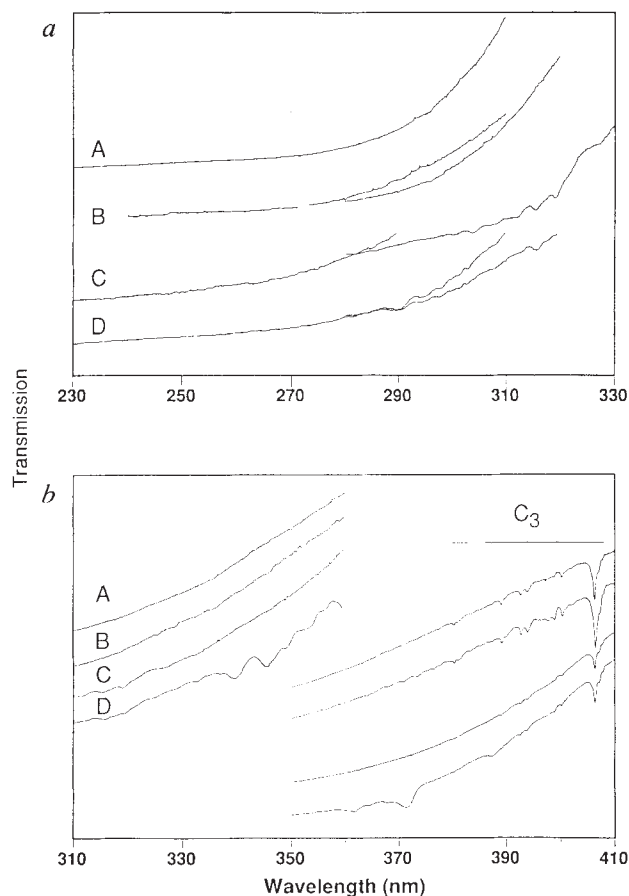


FIG. 1 a and b, Absorption spectra of neon matrices at 5 K in the ultraviolet region measured after codeposition of mass-selected $C_6H_m^-$, $C_8H_m^-$, $C_{10}H_m^-$ or $C_{12}H_m^-$ and neon, with concomitant Hg lamp irradiation (traces A, B, C and D, respectively). Spectra were recorded in sections due to light and detector characteristics, and scattering effects. Fragmentation is indicated by the appearance of the C_3 band system. Selection of $C_{10}H_m^-$ or $C_{12}H_m^-$ results in absorption bands (320–370 nm) similar to those produced by aromatic ring systems.

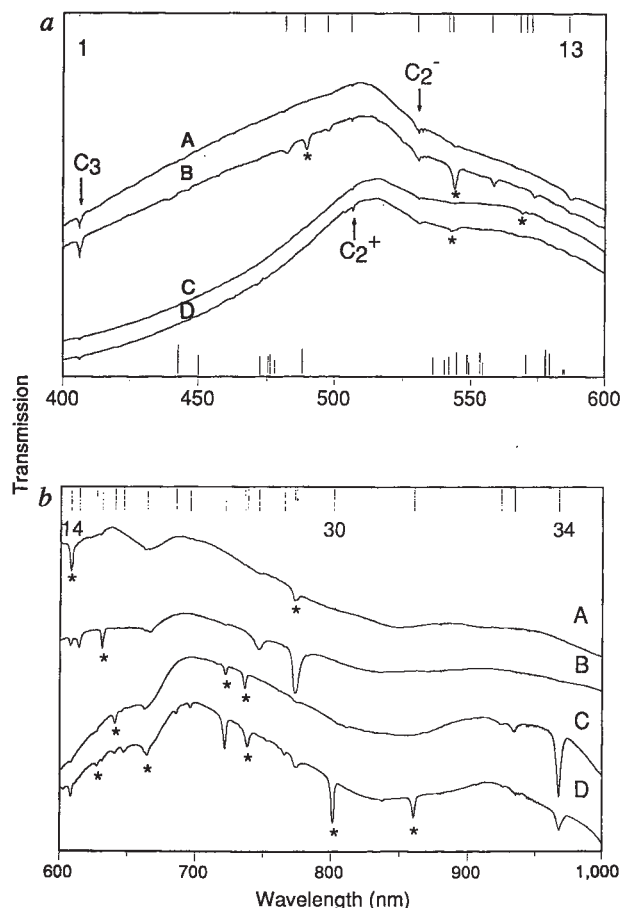


FIG. 2 *a* and *b*, Absorption spectra of neon matrices at 5 K in the range 400–1,000 nm; deposition conditions as Fig. 1. Bands characteristic of C_3 , C_2^+ and C_2^- , indicate that fragmentation processes have taken place. The absorption spectrum resulting from C_8H_m selection (trace B), shows two strong bands, which lie close to DIBs. The latter are shown schematically at the bottom of the figure on a logarithmic scale to equivalent width^{1,6}. The observed bands are numbered (vertical bars) at the top and are listed in Table 1.

In the 600–1,000 nm range the absorption spectra are composed of narrow (for matrix-isolation conditions) and strong bands, which increase in number with the size of the selected ions (Fig. 2*b*). Their bandwidths (full-width half-maximum, FWHM) are 30–80 cm^{-1} . For comparison, the narrowest absorption band we observe in neon matrices is 20 cm^{-1} for C_2^+ (where the peak corresponds to a species in a single type of site in the matrix; more typical is C_3 whose origin band has FWHM of $\sim 50 \text{ cm}^{-1}$). Some of the bands (for example, centred at 607.7 nm) show multiplet structure due to different orientations of the species in the neon matrix environment.

Fragmentation and recombination processes are also signalled in this spectral range; for example, the band at 607.7 nm is seen in all the spectra. In fact, the intensity pattern of the observed bands can be changed with deposition conditions, suggesting that several different matrix-isolated C_nH_m species are responsible for the absorptions. Thus, in the case of experiments where $n=10$ and 12, the ultraviolet absorptions (320–370 nm) are characteristic of one molecule, but the bands in the 600–1,000 nm belong to a mixture of other species.

Bands marked with stars in Fig. 2*a* and *b* correlate with DIB positions. This is seen better in Fig. 3, where the expanded spectrum is compared with the positions of the DIBs taken from the compilation of Herbig and Leka⁶. Wavelength shifts of electronic transitions between neon matrices and gas phase are likely to be $< 100 \text{ cm}^{-1}$, although the direction of shift seems to be

random, as is especially well documented for infrared transitions¹⁵. An extensive list for comparison of neon-matrix and gas data for neutral radicals is not available, but it has been shown that the spectral shifts for the ground state are the least for neon among rare-gas matrices, and are generally $< \pm 1\%$ (ref. 15).

We take the radical C_3 as a reference point; its origin band is shifted 36 cm^{-1} to the blue of the gas-phase value¹⁵. This shift corresponds to $\pm 1.3 \text{ nm}$ at 600 nm, and $\pm 2.3 \text{ nm}$ at 800 nm. Using this as a selection criterion, only DIBs which lie within

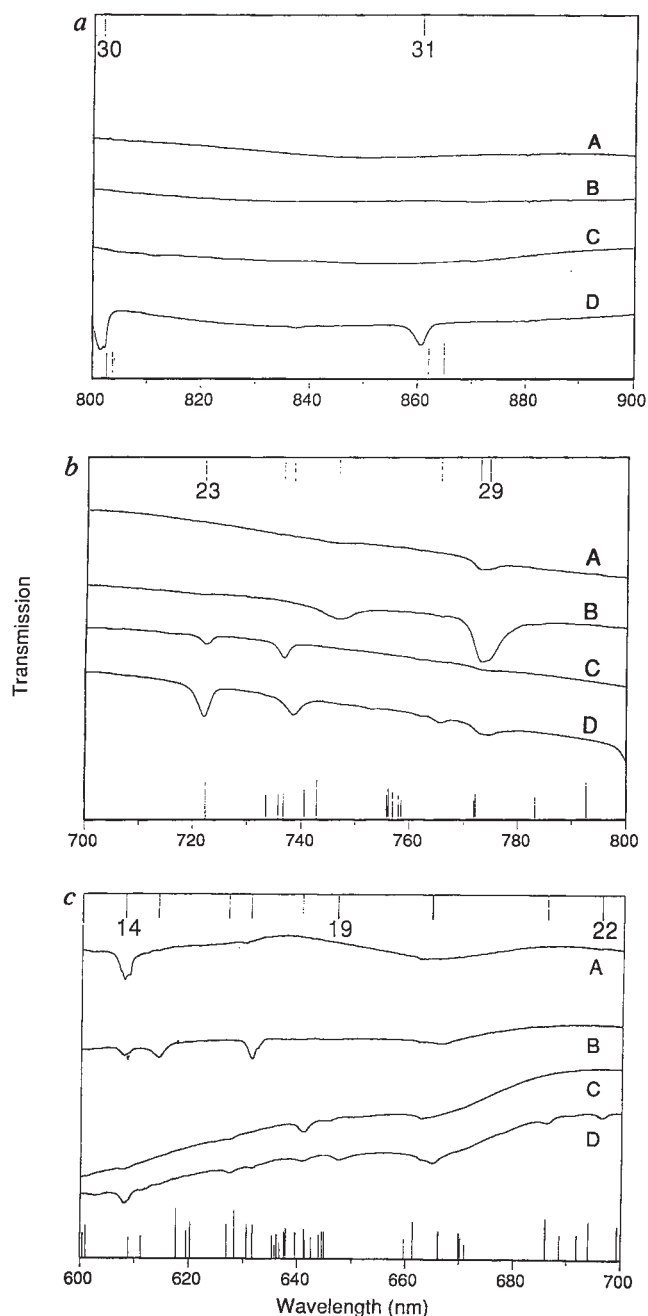


FIG. 3 *a*, *b* and *c*, Absorption spectra of neon matrices at 5 K in the range 600–1,000 nm, deposition conditions as Fig. 1. The 21 observed absorption bands, numbered at the top of each panel, are listed in Table 1. Eleven bands labelled with asterisks lie within $\pm 36 \text{ cm}^{-1}$ (see text) of known DIB positions⁶ are shown at the bottom of each panel by vertical lines with logarithmic scale of their equivalent width.

$\pm 36 \text{ cm}^{-1}$ of our data are listed in Table 1. It should be noted that apart from the matrix shift, an uncertainty also arises because we are using the maxima of the bands in the matrix rather than their zero-phonon line. The latter is only partially resolved on a few of the most intense peaks; for example, the peaks with maxima at 607.7 and 801.9 nm appear to have a zero-phonon line at 608.6 and 802.4 nm respectively. Perhaps the best comparison is a visual one. As shown in Fig. 3, the agreement with some DIBs, which appear spectroscopically 'isolated', is particularly striking.

Bands at 607.7, 772.8 and 967.6 nm dominate the absorption spectra when C_nH_m^- , where $n=6, 8$, or 10, respectively, is mass-selected. The 746.6 nm band appears to be a first member of the progression based on 772.6 nm (with intensity ratio $\sim 1:8$) and the 934.4 nm peak is based on 967.6 nm as origin. These pairs in particular should be good tests of our observations for astrophysical measurements and searches. Bands which do not match in wavelength with reported DIBs are weak in our spectra, except for the moderately intense one at 614.1 nm. (A band at 614 nm appears in a published interstellar spectrum, although there is no mention of its wavelength, or that it is a DIB band¹⁶).

It should also be remarked that in Table 1 relative intensities of the bands are not given, because this can be varied by experimental conditions. Clearly each anion precursor, which appears to be mainly C_nH^- , produces several distinct chemical species as indicated by the number of absorption bands. Fragmentation, recombination and hydrogen atom migration can easily take place during deposition, and will play a role in the final species

distribution. It can be noted here that a similar spectrum as shown for $n=8$ has been obtained by mass-selecting the anion species from phenylacetylene, C_6H_6 , suggesting that some of the hydrogenated carbons produced contain more than one hydrogen atom.

Thus, we propose that the species that produce the observed bands are highly unsaturated hydrocarbon radicals, C_nH_m . One can also exclude the bare carbon molecules, C_n , as no absorption bands in the 600–1,000 nm range have been detected in three different matrix-isolation experiments of carbon vapour^{17–19}.

In the particular case of the C_6H and C_8H radicals, these have been shown to have $^2\Pi$ ground states with acetylenic structures²⁰, and are isoelectronic with tri- and tetra-acetylene cations respectively. The $^2\Pi$ – $^2\Pi$ electronic transitions of the latter are experimentally known in the gas phase, with origin bands at 599.9 nm (C_6H_2^+) and 706.2 nm (C_8H_2^+) (ref. 21). The strongest peaks in the $n=6$ and $n=8$ depositions lie at 607.7 and 772.8 nm and so may well be due to C_6H and C_8H . The electronic spectra of the polyacetylenes C_nH_2 ($n=4, 6, 8, 10$) are also known (their origins lie in the 240–380 nm region)²² but were not detected.

An appealing feature of species such as C_6H and C_8H is that they have an open-shell electronic structure, and so have strong electronic absorptions in the red part of the spectrum (for example, we estimate from our data that the extinction coefficient for the 772.8 nm band is $\epsilon > 10^4 \text{ l mol}^{-1} \text{ cm}^{-1}$). From the astrophysical point of view, partially hydrogenated species are interesting because of the abundance of H atoms in the diffuse clouds and the fact that carbon is a main 'heavy' element. Also, CH and CH^+ are well known species in diffuse clouds²³. Furthermore, in one astrophysical study, it was concluded that the species responsible for some of the interstellar bands (for example, 578 nm—for which we have no corresponding matrix-isolated band in this work) is a neutral gaseous species and the strength of this band depends on the column density of H rather than that of H_2 (ref. 24). Also, C_nH ($n=2$ –6) species have been identified in dense molecular clouds by their rotational spectra²⁵, and C_4H has been detected in absorption in diffuse gas²⁶.

Thus it appears that we may have found a subset of the molecular species which are responsible for some of the DIBs. Other DIBs may well be associated with related unsaturated C_nH_m species. The number of correlations between the absorption bands in the neon matrix and diffuse interstellar bands should serve as a basis for gas-phase studies in the laboratory and for astrophysical measurements. □

TABLE 1 Comparisons of wavelengths

No.	λ_{matrix} (nm)	Spectrum	λ_{DIB} (nm)
1	405.8	C_3 ; A, B, C, D	
2	481.9	B	
3	489.0	B	488.20
4	497.5	B	
5	506.2	C_2^+ ; B, C, D	
6	530.6	C_2^- ; A, B, D	
7	541.9	D	542.00
8	543.5	B , A, D	544.90
9	557.8	B	
10	568.4	C	
11	570.8	C	570.51
12	572.8	B	
13	586.5	A	
14	607.7	A , B, C, D	608.97
15	614.1	B	
16	627.1	D	626.97
17	631.2	B , D	631.80
18	640.7	C , D	639.69 } 641.26 }
19	647.2	D	
20	664.6	D	666.07
21	685.9	D	
22	696.0	D	
23	721.9	D , C	722.40
24	736.4	C	735.75
25	738.3	D	736.70
26	746.6	B	
27	765.5	D	
28	772.8	B , A	772.17
29	774.5	D	
30	801.9	D	802.62
31	860.7	D	862.08
32	924.6	C , D	
33	934.4	C , D	
34	967.6	C , D	

Wavelengths of observed absorption band maxima in a neon matrix ($\lambda_{\text{matrix}} \pm 0.5 \text{ nm}$) and those diffuse interstellar bands (λ_{DIB}) that lie within $\pm 36 \text{ cm}^{-1}$ (see text). A, B, C, D refer to four traces in Figs 1–3. Bold letters indicate trace where peak is dominant.

Received 6 August; accepted 18 October 1993.

- Herbig, G. H. *Astrophys. J.* **196**, 129–160 (1975).
- Krelowski, J. *Astr. Nachr.* **310**, 255–263 (1989).
- Sanner, F., Snell, R. & Vanden Bout, P. *Astrophys. J.* **226**, 460–468 (1978).
- Herbig, G. H. *Astrophys. J.* **331**, 999–1003 (1988).
- Joblin, C., Maillard, J. P., d'Hendecourt, L. & Léger, A. *Nature* **346**, 729–731 (1990).
- Herbig, G. H. & Leka, K. D. *Astrophys. J.* **382**, 193–203 (1991).
- Douglas, A. E. *Nature* **269**, 130–132 (1977).
- Hare, J. P. & Kroto, H. W. *Acc. Chem. Res.* **25**, 106–112 (1992).
- Polycyclic Aromatic Hydrocarbons and Astrophysics* (eds Léger, A., d'Hendecourt, L. & Boccaro, N.) (Reidel, Dordrecht, 1987).
- Salama, F. & Allamandola, L. J. *Astrophys. J.* **395**, 301–306 (1992).
- Salama, F. & Allamandola, L. J. *Nature* **356**, 42–43 (1992).
- Ellis, H. B. Jr & Ellison, G. B. *J. chem. Phys.* **78**, 6541–6558 (1983).
- Fulara, J., Jakobi, M. & Maier, J. P. *Chem. Phys. Lett.* **206**, 203–209 (1993).
- Cheshnovsky, O., Pettiette, C. L. & Smalley, R. E. in *Ion and Cluster Ion Spectroscopy and Structure* (ed. Maier, J. P.) 373–415 (Elsevier, Amsterdam, 1989).
- Jacox, M. E. *J. phys. Chem. Ref. Data* **17**, 269–511 (1988).
- Snow, T. P. *Astrophys. J.* **401**, 775–777 (1992).
- Weltner, W. Jr & McLeod, D. J. *chem. Phys.* **45**, 3096–3105 (1966).
- Krättschmer, W. & Nachtigall, K. in *Polycyclic Aromatic Hydrocarbons and Astrophysics* (eds Léger, A., d'Hendecourt, L. & Boccaro, N.) 75–83 (Reidel, Dordrecht, 1987).
- Szczepanski, J. & Vala, M. *J. phys. Chem.* **95**, 2792–2798 (1991).
- Pauzat, F., Ellinger, Y. & McLean, A. D. *Astrophys. J.* **369**, L13–L16 (1991).
- Allan, M., Kloster-Jensen, E. & Maier, J. P. *Chem. Phys.* **7**, 11–18 (1976).
- Kloster-Jensen, E., Haik, H. J. & Christen, H. *Helv. Chim. Acta* **57**, 1731–1744 (1974).
- Duley, W. W. & Williams, D. A. *Interstellar Chemistry 14* (Academic, London, 1984).
- Herbig, G. H. *Astrophys. J.* **407**, 142–156 (1993).
- Irvine, W. M., Oshishi, M. & Kaifu, N. *Icarus* **91**, 2–6 (1991).
- Bell, M. B., Feldman, P. A. & Matthews, H. E. *Astrophys. J.* **273**, L35–L39 (1983).

ACKNOWLEDGEMENTS. We thank H. W. Kroto for discussion. This work was supported by the Swiss Science Foundation.

Contact electrification induced by monolayer modification of a surface and relation to acid–base interactions

R. G. Horn*, D. T. Smith & A. Grabbe*

Ceramics Division, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, USA

ELECTRICAL charge separation following contact between two materials (contact electrification or the triboelectric effect) is well known to occur between different materials as a consequence of their different electronic structures^{1,2}. Here we show that the phenomenon occurs between two surfaces of the same material if one is coated with a single chemisorbed monolayer. We use the surface force apparatus³ to study contact electrification⁴ and adhesion between two silica surfaces, one coated with an amino-silane. The presence of this monolayer results in significantly enhanced adhesion between the surfaces, owing to electrostatic attraction following contact electrification, in accord with Derjaguin's electrostatic theory of adhesion⁵. At the same time, the observed increase in adhesion is consistent with Fowkes' acid–base model⁶ (in which acid–base interactions between surface groups are considered to be the predominant factor determining adhesion), as the monolayer converts the originally acidic silica surface to a basic (amine-terminated) one. These observations demonstrate a link between acid–base interactions and contact electrification.

Adhesion and contact electrification measurements⁴ were conducted using the surface force apparatus³, with two smooth silica surfaces prepared by the blown-bubble method⁷. One of the surfaces was chemically modified by adsorbing a monolayer of (γ -aminopropyl)dimethylethoxysilane. The method is described by Grabbe⁸, who presents evidence for a stable compact monolayer that is positively charged in aqueous solution, indicating that the silane is covalently bound to the surface with the amine end group facing out.

In situ electrometer measurements⁴ showed that the treated and untreated surfaces were unchanged before they made contact, but acquired charge densities of -2 to -3 mC m⁻² on the bare silica and an equal positive charge on the amine-coated silica following contact in an atmosphere of dry nitrogen. Because of contact charging, a large electrostatic attraction between the bare and the amine-coated silica surfaces was measured (Fig. 1). During separation of the surfaces, a series of partial discharges was detected, clearly visible as breaks in the force–separation curve. These discharges (also observed following silica–mica contacts⁹) occur across the gap between the surfaces, and limit the amount of charge that is fully separated and so measured by the electrometer.

A family of curves in Fig. 1 shows the calculated electrostatic pressure corresponding to different fixed values of surface charge density, based on a simple capacitor model appropriate to the experimental configuration⁹. The measured force is well described by the solid line, which follows a sequence of curves of successively lower charge density connected at the separations where discharges were observed. The charge density required to fit the long-distance tail of the force curve is comparable to the electrometer measurement quoted above. A higher charge density was present when the surfaces were first separated, before the discharges occurred. The fitted value of 6.9 mC m⁻² for separations $D < 1$ μ m corresponds to one electronic charge per 23 nm². This represents $\sim 2\%$ of the ionizable surface groups being charged, because the area per silanol group on the bare

silica as prepared for these experiments, and the area per amine group on the coated silica, are both estimated⁸ to be ~ 0.5 nm².

The work of adhesion required to completely separate the treated and untreated silica surfaces is given by the area between the force curve and the separation axis, which is calculated from Fig. 1 as 3.3 J m⁻². For comparison, the work of adhesion between two identical bare silica surfaces prepared and measured by the same techniques is only 0.1 J m⁻² (ref. 7). Although the enhanced adhesion can certainly be rationalized in terms of the electrostatic description given above, it could also be thought of in other terms. There is a popular view that acid–base interactions across the interface between different materials are a major factor in determining the adhesion between them⁶; this model has been very successful in describing adhesion between glass and polymers in particular, and in guiding the choice of materials or of surface treatments to promote adhesion¹⁰. In the experiment described here one of the surfaces (silica) is acidic due to the presence of dissociable silanol groups¹¹, whereas adsorption of the monolayer has rendered the other surface basic due to its termination with amine groups⁸. The observed polarity of surface charge (silica becoming negative) is consistent with these assignments. Thus the strong adhesion observed between bare silica and amine-coated silica can be considered a manifestation of acid–base adhesion.

Rather than trying to make a distinction between the electrostatic and acid–base models to account for the observed adhesion, it is suggested that there may in fact be a considerable

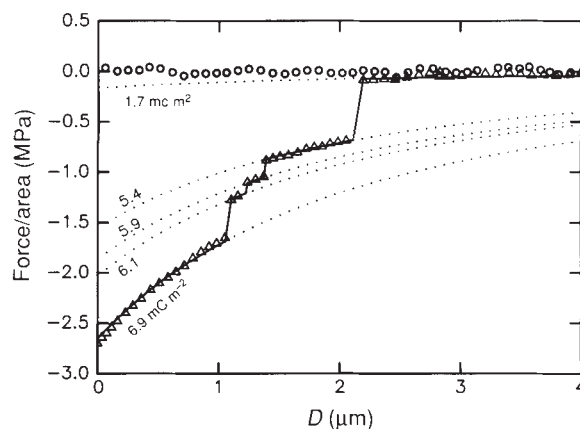


FIG. 1 The force measured during separation from contact between two smooth silica surfaces, one of which is coated with an amine-terminated ($\equiv\text{SiO}-\text{Si}(\text{CH}_3)_2(\text{CH}_2)_3\text{NH}_2$) monolayer (Δ). Negligible force ($\circ$) was measured on first approach of the surfaces—van der Waals attraction (0.005 MPa at 0.01 μ m) is too weak to be seen on this scale. Force F is divided by contact area A_c (1.4×10^{-8} m²) and plotted as a function of distance D between the surfaces. Measurements were made in a surface force apparatus³ in an atmosphere of dry nitrogen. The substantial attractive force displayed here occurs as a result of contact electrification, so that the surfaces bear opposite charges (silica becoming negative) after separation. The Derjaguin approximation, which is commonly used in analysing experiments performed with the surface force apparatus, is not employed here because of the long-range nature of the force and the localized surface charge. The dotted lines show the electrostatic pressure calculated from a simple capacitor model incorporating the effect of grounded electrodes (used for measuring the surface charge⁴) at the outer surfaces of the two silica sheets. Because charge was not measured simultaneously with force, charge densities (given on the figure) have been treated as adjustable parameters in fitting different sections of the data. The data follow the curves if one allows for jumps from one constant-charge curve to another of lower charge density (as shown by the solid line) at certain separations where electrical discharges⁹ were noted.

* Present addresses: Schools of Applied Physics and Chemical Technology, University of South Australia, The Levels, South Australia 5095, Australia (R.G.H.); Sandia National Laboratory, Department 1114, Albuquerque, New Mexico 87185, USA (A.G.).

degree of overlap between the two theories. Conventionally, contact electrification is explained in the physicist's language of electronic states and work functions differing between two materials. The chemist may prefer to speak of acid-base interactions between surfaces, but these also imply a degree of charge rearrangement at the interface, either involving protons (Brønsted acids and bases) or electrons, in the more general Lewis sense. If interacting surfaces are abruptly separated, some fraction of the charge may remain on the 'wrong' material; in other words, acid-base interactions (insofar as they involve an ionic component) may be a precursor to contact electrification. Lee¹² has previously suggested that surface acidity and basicity can be related to surface acceptor and donor states, thus linking the languages of chemistry and physics. Whereas work functions for metals can be measured reliably, they are not easily established for insulators, whose surface states are strongly affected by restructuring, defects and adsorbates. On the other hand, the acid-base nature of insulating surfaces can be readily determined, for example by wetting experiments¹³, and so these characteristics may serve as useful predictors of the sign and magnitude of contact electrification between insulators. Whichever view is preferred, it is clear that the surface treatment

renders the interface sufficiently asymmetric to achieve contact electrification and strong adhesion.

We suggest that this approach may lead to applications in forming exceptionally thin adhesive joints between mating surfaces of materials having the same or similar surface chemistries. □

Received 1 September; accepted 15 October 1993.

1. Harper, W. R. *Contact and Frictional Electrification* (Oxford Univ. Press, 1967).
2. Lowell, J. & Rose-Innes, A. C. *Adv. Phys.* **29**, 947-1023 (1980).
3. Israelachvili, J. N. & Adams, G. E. *J. chem. Soc., Faraday Trans. 1* **74**, 975-1001 (1978).
4. Smith, D. T. *J. Electrostat.* **26**, 291-308 (1991).
5. Derjaguin, B. V., Krotova, N. A. & Smilga, V. P. *Adhesion of Solids* (Consultants Bureau, New York, 1978).
6. Fowkes, F. M. *J. Adhesion Sci. Technol.* **1**, 7-27 (1987).
7. Horn, R. G., Smith, D. T. & Haller, W. *Chem. Phys. Lett.* **162**, 404-408 (1989).
8. Grabbe, A. *Langmuir* **9**, 797-801 (1993).
9. Horn, R. G. & Smith, D. T. *Science* **256**, 362-364 (1992).
10. Mittal, K. L. & Anderson, H. R. (eds) *Acid-Base Interactions: Relevance to Adhesion Science and Technology* (VSP, Utrecht, 1991).
11. Papirer, E. & Balard, H. *J. Adhesion Sci. Technol.* **4**, 357-371 (1990).
12. Lee, L. H. (ed.) *Fundamentals of Adhesion* Ch. 1 & 12 (Plenum, New York, 1990).
13. Fowkes, F. M. *J. Adhesion Sci. Technol.* **4**, 669-691 (1990).

ACKNOWLEDGEMENTS. We thank R. Sharma and the Eastman Kodak Company for their interest and for financial support.

Synchronous changes in atmospheric CH₄ and Greenland climate between 40 and 8 kyr BP

J. Chappellaz*, T. Blunier†, D. Raynaud*, J. M. Barnola*, J. Schwander† & B. Stauffer†

* Laboratoire de Glaciologie et Géophysique de l'Environnement, BP 96, 38402 St Martin d'Hères Cedex, France

† Physikalisches Institut, Sidlerstrasse 5, CH-3012 Bern, Switzerland

ICE-CORE reconstructions of atmospheric methane concentrations for the past 220 kyr have revealed large variations associated with different climatic periods¹⁻⁴. But the phase relationship between climate and methane has been uncertain because of dating uncertainties and the coarse sampling interval of available methane records. Here we present a high-resolution record of atmospheric methane from 40 to 8 kyr ago from the GRIP ice core in Greenland. Our improved resolution and dating allow us to conclude that the large changes in atmospheric methane concentration during the last deglaciation were in phase (± 200 years) with the variations in Greenland climate. Our results confirm the previous observation³ that methane increased to Holocene levels when much of the Northern wetlands was still ice-covered, lending support to the suggestion³ that low-latitude wetlands were responsible for the observed changes. We observe oscillations in methane concentration associated with the warm periods (interstadials) that occurred throughout the glacial period⁵, suggesting that the interstadials were at least hemispheric in their extent. We propose that variations in the hydrological cycle at low latitudes may be responsible for the variations in both methane and Greenland temperature during the interstadials.

Past variations of atmospheric CH₄ concentrations can be reconstructed by analysing the composition of bubbles trapped in polar ice sheets. Air filling the pore space in firn is enclosed when firn sinters to ice. The enclosure occurs at a depth of typically 70 m below the surface in a slow, continuous process. This implies that the air in the bubbles is younger than the surrounding ice and that very fast (for example, seasonal) variations are smoothed by the slow enclosure process.

Several arguments support the reliability of the ice-core reconstruction of CH₄ variations within the experimental uncertainties⁶⁻⁸. Earlier measurements from Dye 3 (Greenland) and Byrd and Vostok (Antarctica) showed that glacial-interglacial climatic changes were accompanied by large variations of the atmospheric CH₄ concentration, and that a similar relationship might also apply to events of shorter duration like the Younger Dryas. There remained questions about the potential leads and lags between CH₄ and climatic signals, and about possible CH₄ changes accompanying the interstadial events of the last ice age in the North Atlantic region. The new GRIP (Greenland Ice-core Project) ice core obtained on the top of the Greenland ice sheet at Summit (72° 34' N, 37° 37' W, altitude 3,230 m) offers the opportunity to answer these questions owing to (1) good time resolution caused by the large accumulation rate (present-day 0.21 m-H₂O yr⁻¹) at Summit, (2) precise dating of the ice⁵, (3) a relatively short age difference between air and surrounding ice (210 yr, compared with 2,500 yr at Vostok under present-day conditions^{9,10}), (4) an excellent ice-core quality and (5) the measurements in progress of various chemical compounds in the same core, which can provide additional information for the interpretation of the CH₄ record.

Measurements of methane concentration have been performed at LGGE using a wet extraction technique followed by gas-chromatographic analyses of the extracted air²⁻⁴. The contamination occurring with this technique has been reduced by replacing the grease sealing of the sample container by an indium seal; contamination averages now 18 ± 26 (2 σ) parts per billion by volume (p.p.b.v.). The analytical precision (calibration + contamination) is on average 37 p.p.b.v. (2 σ), excluding the uncertainty of the standard gas.

Ice samples (83) were analysed every ~10 m between 1,330 and 2,270 m depth. This leads to a mean time resolution ranging from 100-200 yr (upper part) to 600-900 yr (bottom part). The timescale for the ice originates from ref. 11. The age difference between air trapped in the bubbles and the surrounding ice has been investigated carefully at Summit for present-day conditions⁹. For other climatic epochs, it was calculated with a semi-empirical model of firn densification¹⁰, taking into account the temperature dependence of the close-off density¹². The model requires knowledge of the mean annual temperature and the snow accumulation rate at the drill site, both deduced from the GRIP $\delta^{18}\text{O}$ profile^{5,11}. The resulting air-ice age difference ranges from 210 yr in present-day conditions to 750 yr in full glacial

conditions. The duration of the bubble enclosure ranges from 20 to ~100 yr respectively.

The atmospheric CH₄ profile is plotted against depth and gas age, together with the GRIP isotopic record^{5,11} in Fig. 1. The record follows the main features revealed by previous Antarctic and Greenland records: the glacial-interglacial doubling from 350 to 700 p.p.b.v., the high mean concentrations around 30–35 kyr BP and the remarkable oscillation during the deglaciation, which is shown by the GRIP record to be synchronous with the Younger Dryas. Surprisingly, the Younger Dryas CH₄ concentration minima at Vostok and GRIP are almost identical. The air enclosure duration at Vostok is theoretically 900 yr (ref. 10), so a smoothing of the atmospheric Younger Dryas signal, which lasts ~1,000 yr, is expected. As this is not observed, we suggest that the air enclosure happened faster at Vostok than assumed.

New and remarkable in the GRIP core is the observation of six oscillations between 40 and 28 kyr BP which were not depicted previously (by the Vostok profile for example), probably because of the coarse sampling interval in this part of the Vostok record. Each of these GRIP CH₄ oscillations is associated with an isotopic interstadial episode (labelled 5, 6, 7, 8, 10 and 11 in Fig. 1). The interstadials labelled 2, 3, 4 and 9 have a shorter duration (~500 yr) and are associated with less-pronounced CH₄ peaks (no peak for interstadial no. 9). This could be due to the limited depth resolution of methane measurements in this part of the record, which could also be the reason for the differences of the ratios of amplitudes between isotopic and CH₄ oscillations for the interstadials 5 to 11. An additional new feature coming from the GRIP record is the distinct CH₄ minimum in the early part of the Holocene at 8.2 kyr BP, associated with a drop in the $\delta^{18}\text{O}$ record.

The differences between the CH₄ and isotope records are also remarkable. The first difference concerns the start of the deglaciation, where the isotope indicates a first slow temperature increase between 20 and 17 kyr BP, whereas the CH₄ level stays at low values until 16.5 kyr BP. Another difference is in the Bølling-Allerød period (14.45–12.7 kyr BP), when the isotope shows a decreasing trend while CH₄ is increasing. Finally, the start of the Pre-Boreal stage (11.5–10.0 kyr BP) also indicates a slow isotopic increase, when the CH₄ level is already at its maximum. These differences indicate that the slow isotopic changes are not reflected in the CH₄ profile. But we note that during the Bølling-Allerød period, palaeodata reveal a warming trend in eastern North America¹³, suggesting that the cooling pattern in Greenland may not reflect an hemispheric phenomenon.

The question of possible time lags between CH₄ and climate changes can best be addressed by the use of our results covering the deglaciation and the Younger Dryas. The two fast CH₄ and temperature increases at 14.45 and 11.55 kyr BP (ref. 5) and the decreases at 12.7 kyr BP are almost in phase. The observed CH₄ changes precede the isotopic shifts by 100–200 yr, but such differences still lie within the uncertainties linked to the temperature and accumulation rate changes introduced in the densification model and are therefore not significant.

The GRIP CH₄ profile brings new insights to the causes of the methane concentration variations. The scenarios proposed to date involve (1) fluctuations in the global wetland extent, (2) clathrate decomposition, and (3) variations in the atmospheric concentration of the OH radical. Photochemical model studies indicate that the changes in the oxidative capacity of the atmosphere probably amplified the changes in CH₄ concentration during the deglaciation, requiring, in any case a major change in the CH₄ source strength as a primary cause (see ref. 14 for a review). Clathrate outgassing could lead to several abrupt increases of the atmospheric CH₄ level, up to thousands of p.p.b.v., during the deglaciation¹⁵. Such features are not observed in the GRIP record although its time-sampling interval is a factor of ten better than the pre-existing records. Moreover, clathrate decomposition due to warming in permafrost regions¹⁵

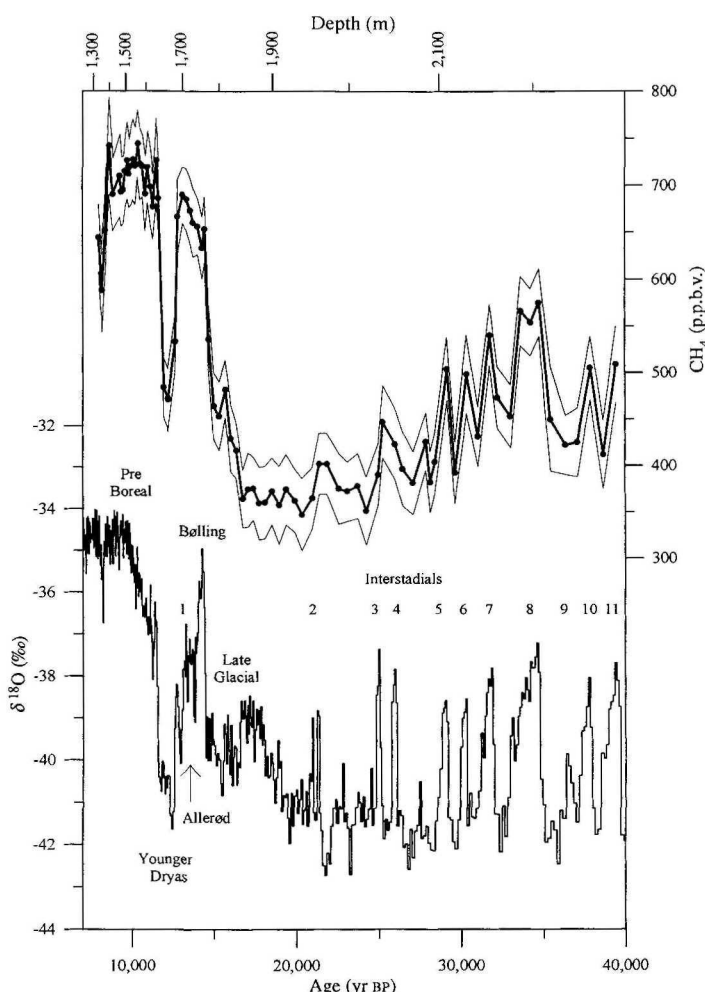


FIG. 1 GRIP ice-core profiles. Top, CH₄ record. The thick line runs through the mean concentration (black dots) and the two accompanying thin lines correspond to the experimental uncertainty (2σ). Bottom, $\delta^{18}\text{O}$ record along 2.2-m sections of the core^{5,11}. The significant climatic events are noted by name or by suggested numbering⁵. The timescale applies to both records. The depth scale applies only to the CH₄ curve (top), because of the difference in age between trapped air and ice at a given depth. The CH₄ data are archived in the Ice Core Data Bank of the World Data Center-A for Paleoclimatology (NOAA, Boulder, Colorado).

would induce a CH₄ lag over the major Greenland warmings, which is not observed in the GRIP record. Although not ruled out, the hypothesis of catastrophic outgassing appears less probable. (It is, however, possible that clathrate outgassing contributed to smaller and slower CH₄ changes in the ice record.) On the other hand, we can quantitatively explain the deglacial CH₄ doubling by the moisture increase on the continents¹⁶. The relative role of high and low latitude wetlands in the control of atmospheric CH₄ levels is still questioned. But we note that CH₄ already reaches interglacial levels during the Bølling, when the Laurentide and Fennoscandinavian ice sheets were still largely extended. The CH₄ increase at the end of the last glaciation therefore leads the development of boreal vegetation and associated wetlands at high latitudes¹⁷, as for example depicted by the comparison with organic acid variations in the GRIP ice core¹⁸. Also, the two CH₄ oscillations (during the Younger Dryas and around 8.2 kyr BP) parallel two episodes of drought well marked in (for example) tropical Africa^{19–21} and Tibet²². These observations reinforce the hypothesis that low-latitude moisture fluctuations play an important role in changes of atmospheric methane concentrations during the deglaciation.

The Greenland isotopic interstadials, as well as the Younger Dryas and 8.2 kyr BP droughts in the northern tropics are thought to result from changes in the formation of North Atlantic deep water (NADW)^{20,23}, possibly linked to instabilities of the Laurentide ice sheet for the glacial episodes²⁴. Therefore, the parallelism between the CH₄ fluctuations and the Greenland interstadial episodes depicted by our record would reflect an intimate connection between the North Atlantic thermohaline circulation and the hydrological cycle over the tropics, if the latter is effectively responsible for the major CH₄ changes. One possible mechanism to generate this connection could be a change in the evaporation rate over the tropical Atlantic, which would affect the advection of warm and saline subtropical water (towards the pole) and hence the NADW formation²⁵, as well as precipitation over land. Such a mechanism could even account for a slight lead of the CH₄ increases over the Greenland abrupt warmings if the changes in evaporation and moisture at lower latitudes would first produce warm and salty water streaming north. A more detailed CH₄ record and progress on air dating can, together with coupled ocean-atmosphere models, help to find the basic mechanisms behind this emerging Greenland climate-methane connection.

As the interhemispheric mixing of the atmosphere is fast compared with the CH₄ changes observed in the GRIP ice core, the latter could be used to establish the link between the ice-core climatic records of the Northern and Southern Hemispheres. So far, the only GRIP CH₄ oscillation seen in the Antarctic ice is the Younger Dryas event in the Vostok profile. In order to synchronize the GRIP and Vostok (dating in ref. 4) Younger Dryas CH₄ oscillations, it is necessary to shift the Vostok oscillation by 1,200 yr (older), either by decreasing the air-ice age difference or by increasing the age of the ice. Uncertainties in these two quantities prevents attribution of the 1,200 yr shift definitely to one or the other. Moreover, re-dating the Vostok deglaciation would become easier with additional CH₄ measurements on this core.

The close relationship between atmospheric CH₄ concentrations and the Greenland isotopic temperature points to a strong biosphere-climate coupling on the secular timescale, and is in agreement with a positive CH₄ forcing on the climate. The magnitude of the CH₄ changes leads however, to a direct radiative impact of (at most) one-tenth of a degree^{3,7} and therefore cannot be responsible for the abrupt shifts of several degrees recorded in Greenland. Additional CH₄ measurements on existing Greenland and Antarctic ice cores are still an important future objective, as they would help to clarify the crucial issues of (1) the phase relationship between the climatic events of the two hemispheres, (2) the interhemispheric CH₄ difference and its variations (that is an additional constraint for the causes of CH₄ changes) and (3) the geographical extent of the Greenland climatic flip-flops observed along the GRIP ice core¹¹. □

Received 2 August; accepted 19 October 1993.

1. Stauffer, B., Lochbrunner, E., Oeschger, H. & Schwander, J. *Nature* **332**, 812–814 (1988).
2. Raynaud, D., Chappellaz, J., Barnola, J. M., Korotkevich, Y. S. & Lorius, C. *Nature* **333**, 655–657 (1988).
3. Chappellaz, J., Barnola, J. M., Raynaud, D., Korotkevich, Y. S. & Lorius, C. *Nature* **345**, 127–131 (1990).
4. Jouzel, J. *et al.* *Nature* **364**, 407–412 (1993).
5. Johnsen, S. J. *et al.* *Nature* **359**, 311–313 (1992).
6. Etheridge, D. M., Pearman, G. I. & Fraser, P. J. *Tellus* **44B**, 282–294 (1992).
7. Raynaud, D. *et al.* *Science* **259**, 926–934 (1993).
8. Raynaud, D. & Chappellaz, J. *The Global Cycle of Atmospheric Methane*, NATO ARW Ser. (ed. Khalil, M. A. K.) (Springer, Berlin, in the press).
9. Schwander, J. *et al.* *J. geophys. Res.* **98**, 2831–2838 (1993).
10. Barnola, J. M., Pimienta, P., Raynaud, D. & Korotkevich, Y. S. *Tellus* **43B**, 83–90 (1991).
11. Dansgaard, W. *et al.* *Nature* **364**, 218–220 (1993).
12. Martinerie, P., Raynaud, D., Etheridge, D., Barnola, J. M. & Mazaudier, D. *Earth planet. Sci. Lett.* **112**, 1–13 (1992).
13. Peteet, D. *et al.* *Quat. Sci. Rev.* (in the press).
14. Thompson, A. M. *Science* **256**, 1157–1165 (1992).
15. Nisbet, E. G. *J. geophys. Res.* **97**, 12859–12867 (1992).
16. Chappellaz, J. A., Fung, I. Y. & Thompson, A. M. *Tellus* **45B**, 228–241 (1993).
17. Adams, J. M., Faure, H. & Petit-Maire, N. *Nature* **355**, 214 (1992).
18. Legrand, M., De Angelis, M. *J. geophys. Res.* (submitted).

19. Street-Perrott, F. A. & Roberts, N. in *Variations in the Global Water Budget* (eds Street-Perrott, A., Beran, M. & Ratcliffe, R.) 331–345 (Reidel, Dordrecht, 1983).
20. Street-Perrott, F. A. & Perrott, R. A. *Nature* **343**, 607–612 (1990).
21. Gasse, F. & Fontes, J. C. in *The Last Deglaciation: Absolute and Radiocarbon Chronologies*, NATO ASI Ser. I2 (eds Bard, E. & Broecker, W. S.) 295–325 (Springer, Berlin, 1992).
22. Van Campo, E. & Gasse, F. *Quat. Res.* **39**, 300–313 (1993).
23. Broecker, W. S. *Oceanography* **4**, 79–89 (1991).
24. Bond, G. *et al.* *Nature* **360**, 245–249 (1992).
25. Duplessy, J. C. *et al.* *Nature* **358**, 485–488 (1992).

ACKNOWLEDGEMENTS. This work is a contribution to the Greenland Ice-core Project (GRIP) which was coordinated and supported by the European Science Foundation. We thank the national funding organizations in Belgium, Denmark, France, Germany, Iceland, Italy, Switzerland and the UK, together with the XII Directorate of CEC, the CNRS Programme de l'Environnement (PACB) and the Fondation de France. We also thank all GRIP participants, M. Legrand and D. Peteet for the communication of unpublished data, R. Delmas and C. Lorius for corrections to the manuscript, and A. Watson and L. Labeyrie for discussions.

Marine barite as a monitor of seawater strontium isotope composition

A. Paytan, M. Kastner, E. E. Martin, J. D. Macdougall & T. Herbert

Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093, USA

THE strontium isotope ratio in sea water is influenced by climate, tectonics, weathering and hydrothermal activity at ocean ridges^{1–4}. Its evolution through time, determined primarily by measuring the strontium isotope composition of marine carbonates, holds information about variations in these processes, and is also useful for stratigraphic correlation and dating^{5–7}. Carbonates are absent from some marine sediments such as siliceous oozes and red clays, and can be significantly diagenetically altered in others, especially in Eocene and older sediments. Here we show that marine barite is an effective alternative monitor of seawater ⁸⁷Sr/⁸⁶Sr. We find that microcrystals of marine barite separated from Holocene Pacific, Atlantic and Indian Ocean sediments all record the modern seawater ⁸⁷Sr/⁸⁶Sr value. Moreover, the ⁸⁷Sr/⁸⁶Sr of barite from 25 sediment samples spanning the past 35 Myr falls within the range of published data for carbonates over this time period. We conclude that marine barite reliably records both present and past variations in seawater strontium isotope composition.

The residence time of dissolved Sr in the oceans is 2–3 Myr, several orders of magnitude longer than the mixing time (~1,500 yr), and thus the Sr isotope composition is uniform throughout the world's oceans. Based primarily on measurements of carbonates, the seawater Sr isotope composition has been determined for the past 500 Myr of geological history⁸. The record is known in considerable detail for the Tertiary period^{9–11} but, although the general features are known, becomes progressively uncertain further back in the geological record.

A limitation in using seawater Sr isotope composition for palaeoceanographic research is the availability and integrity of the recording material. Calcite is known to undergo recrystallization on burial, changing its chemical and/or isotope composition^{12–15}. Furthermore, almost 50% of the ocean floor is covered with non-calcareous sediment. Therefore, identifying other phases that accurately record seawater Sr isotope ratio is an important goal. Marine barite, a widespread component of pelagic sediments^{16,17}, may be such a phase. Within the rather large uncertainty limits available in the late 1960s, Goldberg *et al.*¹⁸ showed that barite in surface sediments has the same Sr isotope composition as sea water.

Marine barite, in the form of microcrystals or aggregates of 0.5–5 µm size, is a ubiquitous minor phase of oceanic particulate matter and pelagic sediments^{19,21}. It is particularly abundant in

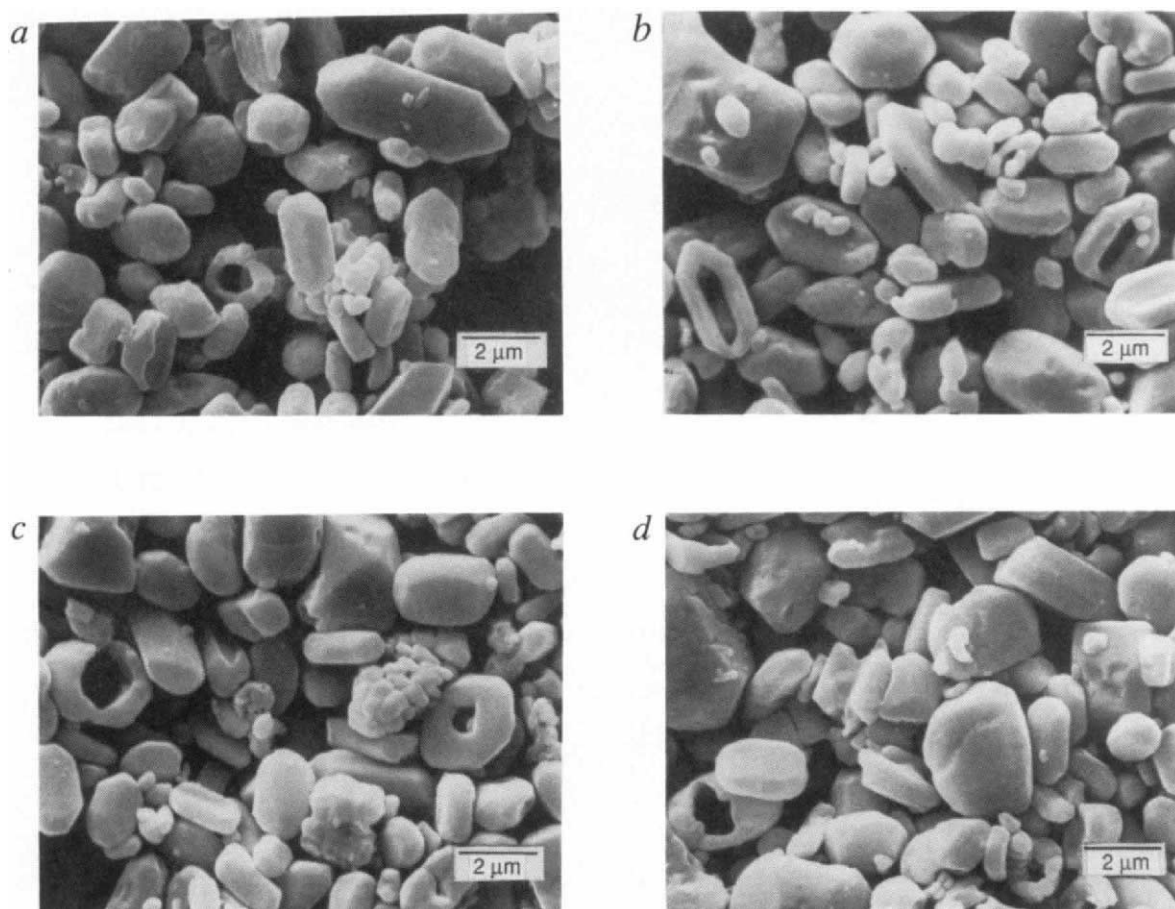


FIG. 1 SEM micrographs of separated marine barite: a, Holocene barite from top of core PLDS 77, 1.03° N, 119.55° W; b, Late Miocene barite from sample 575B-4-6 at 38 m depth, 9.6 Myr old; c, Early Miocene barite from sample 574C-11-3 at 93.5 m depth, 20.4 Myr old; d, Oligocene barite from sample 574C-24-4 at 419 m depth, 29.1 Myr old.

METHODS. Barite separation method modified from Church²¹ as follows: Approximately 10 g (dry weight) of a marine sediment sample is dispersed in distilled water and the carbonate removed by 4 M acetic acid. The residue is thoroughly washed with distilled water and centrifuged to remove all the leachant; this washing is repeated after each of the following leaching steps. To oxidize organic matter the sample is treated with warm (~50 °C) sodium hypochlorite (5 wt%), slightly acidified with HCl to produce Cl₂ gas). The washed residue is leached with 0.02 M hydroxylamine hydrochloride in 25 vol% acetic acid to remove transition-metal oxyhydroxides. The residue is passed through a 30 µm sieve and digested overnight in 50 ml 2:1 mixture of HNO₃ (1 M) and concentrated HF, to remove silicates. The leachant is decanted and the sample leached overnight with a mixture of 1:1 HNO₃ (1 M) and concentrated HF. After a third leach with a mixture of 1:2 HNO₃ (1 M) and concentra-

ted HF the sample is treated with 1:1 saturated AlCl₃ and HNO₃ (1 M) for 1 h at 90 °C to dissolve fluorides. The sample is then washed thoroughly and the residue is ashed in a muffle furnace at 700 °C to remove any refractory organic residue. The sample is analysed with XRD and SEM-EDS, and any of the above treatments is repeated if necessary. A Sr spike which was added to every step of the separation, indicated that no reprecipitation of barite or isotopic redistribution occurs during this separation routine. The residues from this leaching method, a few milligrams of barite, contains minor amounts of Ti oxides and occasionally additional highly refractory minerals. However, because of the low Sr concentration of the refractory minerals and the very high Sr concentration of barite, even if some of this material dissolves, in most cases it should not affect the barite Sr isotope composition. The barite from these samples could be preferentially dissolved by chelation with cation exchange resin (AG-50X8; ref. 21), as used in this study, by diethylenetriaminepenta-acetic acid (DTPA) (ref. 38) or by carbonate-for-sulphate replacement with Na₂CO₃ (ref. 40). A sample of ~0.1 mg of barite is processed for Sr separation using standard ion-exchange techniques.

regions of high biological productivity. In equatorial upwelling zones marine barite reaches concentrations of 1–2 wt% on a carbonate and/or opal-free basis; in the Southern Ocean in biogenic siliceous sediments it reaches concentrations of ~1 wt% on an opal-free basis (for example, core RC-13-256 at 53° S, 0° W) and it has also been detected in red clays. The exact mechanism of barite formation is unknown. There are indications that its precipitation is mediated by biological activity in the upper 2,000 m of the water column, apparently in micro-environments containing decaying organic matter and especially associated with biogenic siliceous remains^{20,22}. Some barite dissolves at depth in the water column, but due to its low solubility and to the fact that most of it is transported to the sea floor within faecal pellets, much of the barite accumulates in the sediment^{21,23,24}. In most pelagic sediments the pore water rapidly becomes saturated with barite, hence most of the barite is preserved in oxic sedi-

ments. Because barite is a rather insoluble mineral it should resist diagenetic exchange after burial. The presence of morphologically and geochemically unaltered marine barite in calcareous sediments with clear indications of carbonate diagenesis (for example, DSDP Leg 85 Site 574, at >250 m burial depth²⁵), supports this suggestion. Minor amounts of particulate barite also form in submarine hydrothermal systems^{26,27}, and barite probably dissolves in sulphate-reducing sediments and re-precipitates at the oxic–anoxic boundaries of such sediments²⁸. These hydrothermal and diagenetic barites may have Sr isotope ratios which are different from contemporaneous sea water, but they are morphologically distinguishable^{26,28} and should not be used indiscriminately for chemical palaeoceanography.

Separation of a pure barite fraction for Sr isotope analyses is difficult because of its relatively low abundance and small grain size. X-ray diffraction (XRD) and scanning electron microscopy

TABLE 1 Measured strontium isotope composition of marine barites

(a) From box core tops

Sample core no.	Lat.	Location Long.	Water depth (m)	Age (^{14}C yr)	Weight% CaCO_3	$^{87}\text{Sr}/^{86}\text{Sr}^*$
Pacific Ocean						
ERDC 77	4.51° N,	156.03° E	3,585	6,495 $\pm$ 100	87.2	0.709165
ERDC 125	0.01° S,	160.59° E	3,368	3,410 $\pm$ 70	82.8	0.709179
ERDC 135	0.52° N,	160.59° E	3,509	4,540 $\pm$ 80	82.5	0.709161
ERDC 139	1.21° N,	162.23° E	4,118	5,590 $\pm$ 110	76.4	0.709182
PLDS 68	1.01° N,	105.29° W	3,650	4,050 $\pm$ 80	72.5	0.709161
PLDS 70	1.03° N,	107.12° W	3,694	4,130 $\pm$ 60	73.4	0.709183
PLDS 77	1.03° N,	119.55° W	4,366	7,530 $\pm$ 215	68.0	0.709166
PLDS 81	1.01° N,	124.32° W	4,771	6,580 $\pm$ 275	47.2	0.709159
PLDS 85	0.58° N,	128.27° W	4,385	4,850 $\pm$ 95	68.0	0.709176
PLDS 107	6.09° N,	138.16° W	4,849	13,710 $\pm$ 690	34.5	0.709185
Atlantic Ocean						
INMD 110	10.02° S,	13.23° W	1,959	4,890 $\pm$ 160	89.5	0.709158
INMD 113	15.15° S,	14.58° W	3,471	4,895 $\pm$ 85	86.3	0.709169
INMD 125	29.40° S,	33.02° W	3,535	4,970 $\pm$ 95	91.4	0.709179
Indian Ocean						
LSDA 107	05.40° S,	70.17° E	3,960	Quaternary	89.7	0.709162
LSDA 106	05.34° S,	63.43° E	4,090	Quaternary	82.3	0.709170
LSDA 107	05.26° S,	59.15° E	4,010	Quaternary	85.9	0.709169

(b) From sediment cores, DSDP Leg 85, Sites 573–575

Site-core-section	Depth (m)	Age (Myr) [†]	$^{87}\text{Sr}/^{86}\text{Sr}^*$
Core tops [‡]	0.00–0.02	0	0.709170
573-1-1	0.8	0.05	0.709187
573-4-3	25.5	1.7	0.709081
573-6-5	45.8	3.2	0.709021
574-4-5	30.4	4.67	0.708981
573-9-6	75.2	4.7	0.708992
573-13-3	105.0	6.0	0.708918
575B-4-3	33.0	7.9	0.708891
Duplicate [§]			0.708932
Duplicate [§]			0.708911
575B-4-6	38.0	9.6	0.708885
575B-5-3	43.0	10.7	0.708890
575B-5-6	47.0	11.3	0.708842
575B-6-5	55.0	12.0	0.708830
Duplicate [§]			0.708836
575B-7-2	57.5	12.3	0.708834
575B-9-2	76.0	13.8	0.708824
575B-11-2	93.5	15.0	0.708792
575B-12-2	101.5	15.7	0.708755
575B-14-4	118.5	16.6	0.708692
574C-5-4	238.5	18.7	0.708561
Duplicate			0.708541
574C-11-3	293.5	20.4	0.708421
574C-13-4	314.5	21.5	0.708360
574C-17-1	347.0	23.4	0.708244
574C-17-5	353.5	24.6	0.708188
574C-19-2	367.5	25.5	0.708152
574C-20-2	378.0	27.0	0.708130
574C-22-3	398.5	28.5	0.708110
Duplicate [§]			0.708113
574C-24-4	419.0	30.9	0.708034
574C-27-2	443.5	32.9	0.707965
574C-31-2	481.0	33.9	0.707926
574C-33-2	498.5	35.5	0.707864
Duplicate			0.707855

* The 2σ error in run uncertainties are all smaller than the external precision of ± 0.000024 . The modern seawater value (NASS-2) measured in the Scripps Institution of Oceanography (SIO) laboratory is 0.709175 ± 24 . All ratios are normalized to $^{86}\text{Sr}/^{88}\text{Sr} = 0.1194$.
 The $^{86}\text{Sr}/^{88}\text{Sr}$ ratio of the NBS 987 standard measured in the SIO lab is 0.710260 ± 24 .

[†] Based on biostratigraphy³² and the timescale of Berggren *et al.*³³.

[‡] Mean value of 16 core top samples from Table 1a.

[§] This duplicate is a separate mass spectrometric run of the same sample.

^{||} This duplicate includes both Sr fraction separation and mass spectrometry.

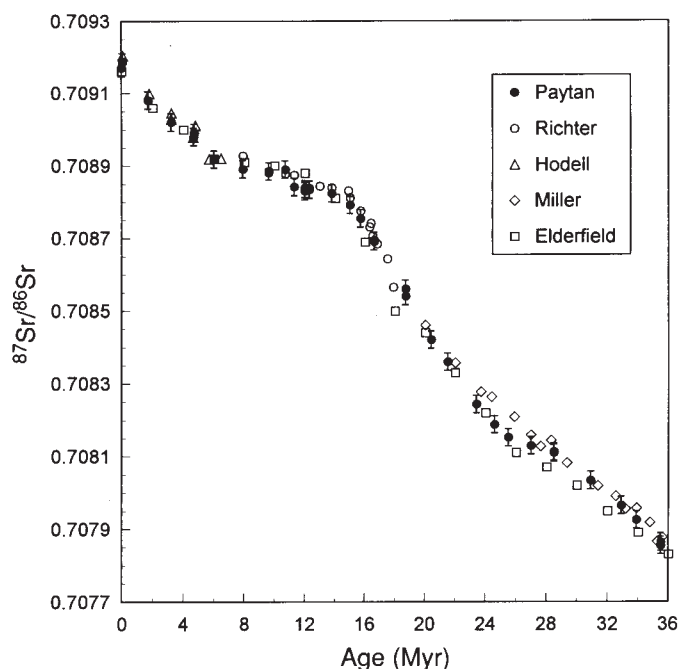


FIG. 2 Seawater Sr isotope composition over the past 35 Myr. Data sources as follows: Paytan, marine barite, this work; Richter, seawater $^{87}\text{Sr}/^{86}\text{Sr}$ ratio at DSDP Hole 575B (7.9–16.6 Myr old), using carbonates (Richter and DePaolo³⁴); Hodell, Sr isotope ratios of planktonic foraminifera of similar ages (0–8 Myr old) to the marine barites (Hodell *et al.*³); Miller, Sr isotope ratios of planktonic foraminifera of similar ages (20–36 Myr old) to the marine barites (Miller *et al.*⁵; Elderfield, Average seawater Sr isotope composition for 2-Myr segments (Elderfield³⁹). All Sr isotope ratios measured in other laboratories were normalized to the value for the NBS 987 standard, 0.710260 and are plotted with respect to the chronology by Berggren *et al.*³³. Symbols for marine barite include the error bars of $\pm 24 \times 10^{-6}$.

with energy-dispersive spectroscopy (SEM-EDS) analysis of barite separated using published methods^{21,29} revealed the presence of admixed silicates, oxides and transition-metal oxyhydroxides. Therefore for this study we developed and employed a considerably modified and more rigorous leaching method than the one used by Church²¹, described in the legend to Fig. 1. Barite was separated from 16 ^{14}C -dated core tops from areas of high productivity in the Pacific, Atlantic and Indian oceans³⁰, and from cores from three Equatorial Pacific sites at 133°W containing older sediment samples (ages of 0.05–35.3 Myr, from DSDP Leg 85 Sites 573, 574 and 575B). Barite from representative samples is shown in Fig. 1. These latter samples represent a considerable range of productivity and depositional environments within the Equatorial Pacific region, from 4°N to 2°S palaeolatitudes. Sedimentation rates vary from 0.4 to $>3\text{ cm kyr}^{-1}$ and carbonate content from 40 to 90 wt%. The Sr/Ca ratios of the calcareous sediments and the dissolved Sr concentrations in pore fluids of $>250\text{-m}$ burial depth at Site 574 indicate that 30–50% of the carbonate has recrystallized²², and that the interstitial-water Sr isotope ratios are significantly different from the contemporaneous seawater values^{25,31}.

Values of Sr isotope ratios are given in Table 1; samples separated from core tops all have values close to that of present-day sea water (0.709175, as measured in our laboratory). Data from older samples separated from the DSDP sediments are plotted against age using biostratigraphy³² and the chronology of Berggren *et al.*³³ in Fig. 2. Clearly, the barite data agree very well with the published Sr isotope data for carbonates in this age interval. In detail, however, there may be small discrepancies between the Sr isotope compositions of carbonates and separ-

ated barite. In Fig. 3 we compare data for carbonates with ages 7.9–16.6 Myr (reported by Richter and DePaolo³⁴) with our measurements on barite separated from the same cores and intervals. This comparison shows that the $^{87}\text{Sr}/^{86}\text{Sr}$ ratios are frequently the same for barites and coexisting calcite, but, where there is a difference, the values for barite tend to be marginally lower than those of the calcite. The discrepancy is small and it is possibly due to uncorrected inter-laboratory bias. A similar effect has, however, been found in several other coexisting carbonate and barite samples measured in our laboratory by Martin³⁵. Based on these data, barite appears to be as reliable a monitor of seawater Sr isotope composition as carbonate over the past 35 Myr. Crystal morphology of the barite over this time interval shows no noticeable change (Fig. 1) and the barite's Sr isotope composition has not re-equilibrated with that of the pore fluid. Both observations suggest that these pelagic barites have not significantly recrystallized over the past 35 Myr. Thus marine barite may be useful for Sr-isotope stratigraphy and correlation in epochs where the seawater Sr curve is known and has a slope, especially in periods or marine environments where the calcium carbonate has undergone diagenetic alteration or is absent, such as in siliceous sediments or red clays.

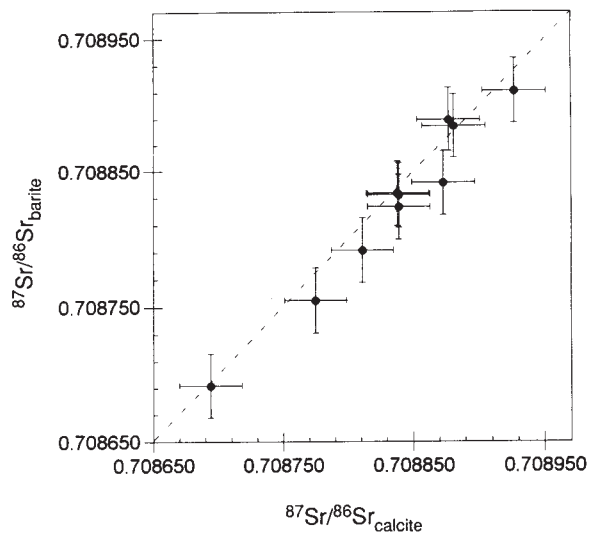


FIG. 3 Sr isotope ratios of marine barites plotted against Sr isotope ratios of marine carbonates (from Richter and DePaolo³⁴) from the same cores and intervals, DSDP 575B (7.9–16.6 Myr old).

Barite also has the potential to be a reliable indicator of other chemical and isotope characteristics of sea water. For example, S isotopes can be measured using a few milligrams of barite, which may lead to a much-improved record of the seawater S isotope composition. Until now, S isotope measurements have been made mostly on evaporite sulphates, which are sporadically distributed in the geologic record. Barite also incorporates a number of minor and trace elements of oceanographic interest, for example Ca, Ra, U, Pb and the rare-earth elements²¹, although at concentrations much lower than that of Sr. Thus in this single phase, it may be possible to monitor for example: (1) the seawater Sr/Ca ratio, (important both for evaluating the degree of carbonate diagenesis¹² and for understanding and modelling the seawater Sr isotope curve with respect to the sinks, sources and size of the oceanic Sr reservoir^{1,2,36}), (2) the seawater Ba/Ca ratio, (important for palaeo-alkalinity studies³⁷) and (3), the Nd isotope ratios, (important for palaeocirculation studies³⁸). □

Received 23 June; accepted 19 October 1993.

- Brass, G. W. *Geochim. cosmochim. Acta* **40**, 721–730 (1976).
- Raymo, M. E., Ruddiman, W. F. & Froelich, P. N. *Geology* **16**, 649–653 (1988).
- Hodell, D. H., Mueller, P. H., McKenzie, J. A. & Mead, G. E. *Earth planet. Sci. Lett.* **92**, 165–178 (1989).
- Richter, F. M., Rowley, D. B. & DePaolo, D. J. *Earth planet. Sci. Lett.* **109**, 11–23 (1992).
- Miller, K. G., Feigenson, M. D., Kent, D. V. & Orson, R. K. *Paleoceanography* **3**, 223–233 (1988).
- Miller, K. G., Feigenson, M. D., Wright, J. D. & Clement, B. M. *Paleoceanography* **6**, 33–52 (1991).
- Hess, J., Scott, L. D., Bender, M. L., Kennett, J. P. & Schilling, J.-G. *Paleoceanography* **4**, 655–679 (1989).
- Burke, W. H. *et al. Geology* **10**, 516–519 (1982).
- DePaolo, D. J. & Ingram, B. L. *Science* **227**, 938–941 (1985).
- Palmer, M. R. & Elderfield, H. *Nature* **314**, 526–528 (1985).
- Hess, J., Bender, M. L. & Schilling, J. G. *Science* **231**, 979–984 (1986).
- Baker, P. A., Gieskes, J. M. & Elderfield, H. *J. sedim. Petrol.* **52**, 71–82 (1982).
- Killingley, J. S. *Nature* **301**, 594–597 (1983).
- Elderfield, H. & Gieskes, J. M. *Nature* **300**, 493–497 (1982).
- Gieskes, J. M., Elderfield, H. & Palmer, M. R. *Earth planet. Sci. Lett.* **77**, 229–235 (1986).
- Arrhenius, G. *The Sea Vol. 3* (Wiley-Interscience, New York, 1963).
- Gurevich, Y. G., Bogdanov, Y. A. & Lisitsyn, A. P. *Geochem. Int.* **15**, 28–42 (1978).
- Goldberg, E. D., Somayajulu, B. L. K., Galloway, J., Kaplan, I. R. & Faure, G. *Geochim. cosmochim. Acta* **33**, 287–289 (1969).
- Dehairs, F., Chesseelet, R. & Jedwab, J. *Earth planet. Sci. Lett.* **49**, 528–550 (1980).
- Bishop, J. K. B. *Nature* **332**, 341–343 (1988).
- Church, T. M. thesis, Univ. California at San Diego (1970).
- Dehairs, F. *et al. Globi Biogeochem. Cycles* **4**, 85–102 (1990).
- Church, T. M. & Wolgemuth, K. *Earth planet. Sci. Lett.* **15**, 35–44 (1972).
- Dymond, J., Suess, E. & Lyle, M. *Paleoceanography* **7**, 163–181 (1992).
- Stout, P. M. thesis, Univ. California at San Diego (1985).
- Feely, R. A. *et al. J. geophys. Res.* **92**, 11343–11363 (1987).
- Feely, R. A., Geiselman, T. L., Baker, E. T. & Massoth, G. J. *J. geophys. Res.* **95**, 12855–12873 (1990).
- Dean, W. E. & Schreiber, B. C. *Init. Rep. DSDP Log* **41**, 915–931 (1977).
- Borole, D. V. & Somayajulu, B. K. *Mar. Chem.* **5**, 291–296 (1977).
- Berger, W. H. & Killingley, J. S. *Mar. Geol.* **45**, 93–125 (1982).
- Baker, P. A. *et al. Earth planet. Sci. Lett.* **105**, 522–533 (1991).
- Barron, J. *Centenary of Japanese Micropaleontology* (Terra Scientific, Tokyo, 1992).
- Berggren, W. A., Kent, D. V. & VanCouvering, J. A. *The Chronology of the Geological Record* (Blackwell, Oxford, 1985).
- Richter, F. M. & DePaolo, D. J. *Earth planet. Sci. Lett.* **90**, 382–394 (1988).
- Martin, E. E. thesis, Univ. California at San Diego (1993).
- Graham, D. W., Bender, M. L., Williams, D. F. & Keigwin, L. D. *Jr Geochim. cosmochim. Acta* **46**, 1281–1292 (1988).
- Lea, D. W. thesis, Woods Hole Oceanogr. Inst., Massachusetts Inst. of Tech. (1990).
- Piepgas, D. J. & Wasserburg, G. J. *Earth planet. Sci. Lett.* **50**, 128–138 (1980).
- Elderfield, H. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **57**, 71–90 (1986).
- Breit, G. N., Simmons, E. C. & Goldhaber, M. B. *Chem. Geol.* **52**, 333–336 (1985).

ACKNOWLEDGEMENTS. We thank G. W. Lugmair and C. S. MacIsaac for guidance in the laboratory. This work was supported by the US National Science Foundation.

Recovery of strain rate variation from inversion of subsidence data

Nicky White

Bullard Laboratories, Madingley Rise, Madingley Road, Cambridge CB3 0EZ, UK

DESPITE considerable advances over the past 20 years in our understanding of continental lithospheric deformation (see, for example, refs 1 and 2), we know little about the detailed spatial and temporal variations in strain rate at geologically relevant scales ($\sim 10^2$ – 10^4 km and ~ 1 –100 million years). Such measurements are needed to constrain dynamic models of lithospheric deformation³. Although it is not yet obvious how detailed strain-rate information can be obtained from mountain belts, subsidence data from sedimentary basins should contain information about at least the vertical component of the strain-rate tensor. Here I show that inverse theory can be used to calculate the temporal strain rate variation needed to fit observed subsidence patterns in extensional sedimentary basins. No *a priori* information concerning either the duration of rifting or the total strain is required. Inverse calculations produce good fits to synthetic data and to subsidence data from the Dolomites of northern Italy. Calculated strain rates are not constant but vary between 10^{-18} and 10^{-15} s⁻¹. Error analysis shows that these results are significant and well resolved. All of the stratigraphic sections studied exhibit a minimum in strain rate during extension—a feature predicted by models of lithosphere deformation by power-law creep.

Strain rate is generally calculated using two very different approaches. Plate tectonic considerations based on sea-floor spreading, mountain building and basin formation yield crude estimates with little spatial or temporal resolution^{4–7}. More recently, geodesy and earthquake seismology have become important tools in estimating short-term spatial strain rate variation in tectonically active regions^{8–10}. Unfortunately, the time period over which strain rates can be measured (10^{-4} Myr) is extremely short compared with the likely duration of deformation (10^1 – 10^2 Myr). Here, a new method for calculating strain rate variation on geologically useful time and length scales is described and applied.

It is generally agreed that the lithospheric stretching model¹¹ (or minor variants thereof) accounts satisfactorily for the general kinematic evolution of most extensional sedimentary basins.

The model assumes that the lithosphere stretches either instantaneously or over a known period of time with constant strain rate¹². But where high-quality subsidence data are available, it is clear that the observed syn-rift subsidence is more variable than that predicted by the constant-strain-rate model^{13,14}. It is important to note that observed syn-rift subsidence variations usually occur over periods of up to ~ 20 Myr. The current uncertainty in calibration of the Phanerozoic timescale is $\sim 2\%$ of the age (for example, ± 5 Myr for the Triassic period) and so inaccuracies in the timescale cannot in general explain these departures from model predictions. Syn-rift subsidence patterns also vary from basin to basin and are rarely the same over areas greater than about $\sim 10^4$ km². Thus global sea-level change cannot be invoked. The documented changes in rate of syn-rift subsidence are explained more easily if the lithospheric strain rate is permitted to vary as rifting proceeds. A method for extracting strain rate variation directly from subsidence data will now be outlined.

For simplicity, the calculations presented here are one-dimensional, although a three-dimensional generalization is feasible⁹. This assumption prescribes a strain-rate tensor for the deformation zone of which only two elements are non-zero: if x and y are the horizontal coordinates and z the vertical coordinate (measured from base of lithosphere) then

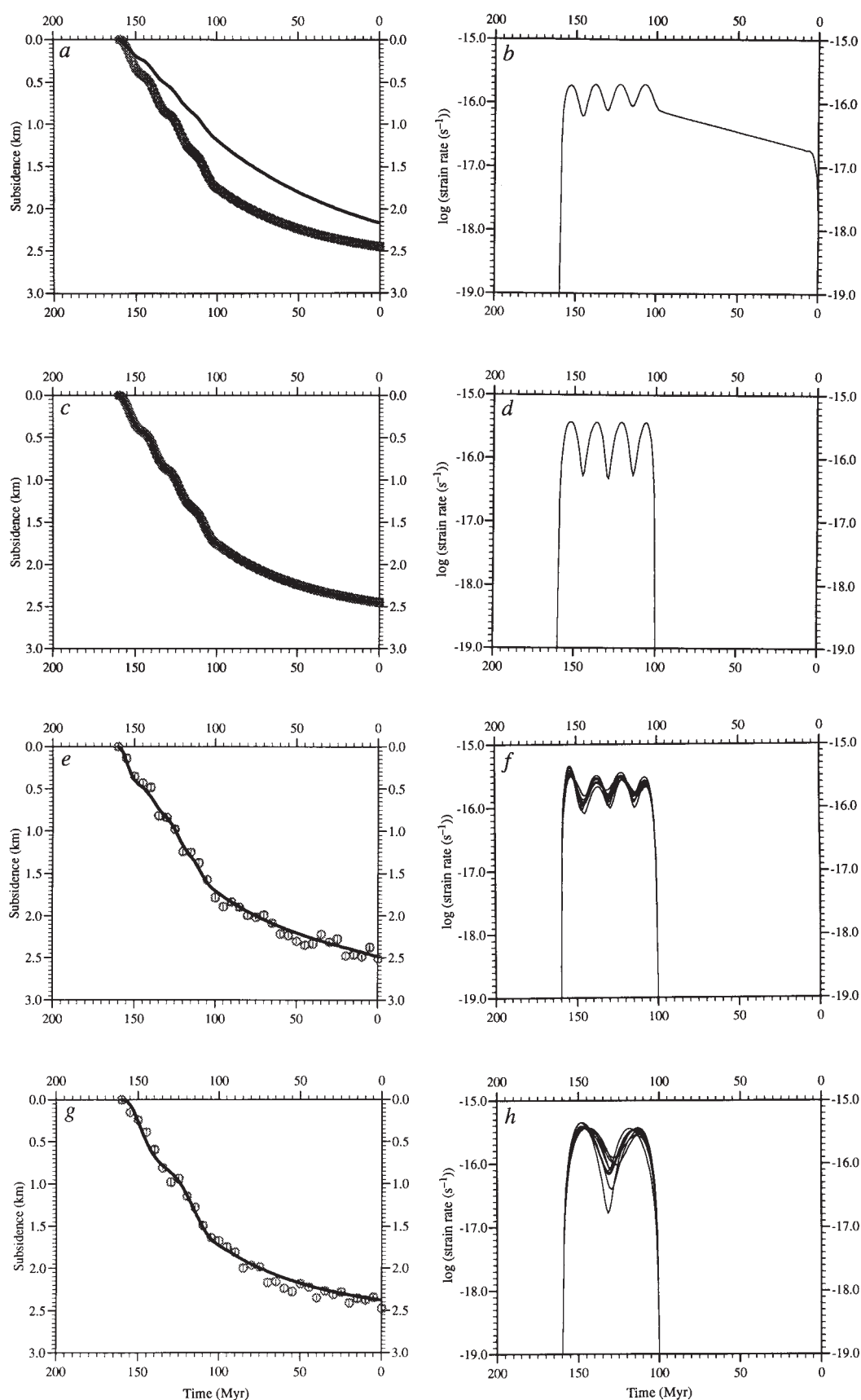
$$\dot{\epsilon}_{zz} = -\dot{\epsilon}_{xx} = G(t)$$

When the uniform stretching model is usually applied, the duration of stretching is assumed to be known from geological constraints and $G(t)$ is set to be constant^{13,14}. Theoretical subsidence curves are then generated as a function of β , the stretching factor (that is, total strain). These curves are fitted to the data in order

TABLE 1 Parameters used in text¹⁹

Symbol	Parameter	Value
a	Lithospheric thickness	125 km
t_c	Crustal thickness	30 km
ρ_w	Seawater density	1.0 g cm^{-3}
ρ_c	Crust density (at 0°C)	2.8 g cm^{-3}
ρ_m	Mantle density (at 0°C)	3.33 g cm^{-3}
ρ_a	Asthenosphere density (at $1,333^\circ\text{C}$)	3.2 g cm^{-3}
α	Thermal expansion coefficient	$3.28 \times 10^{-5} \text{ }^\circ\text{C}^{-1}$
β	Uniform stretching factor	
Δt	Duration of stretching	
T	Temperature	
G	Strain rate	

FIG. 1 Inverse modelling for synthetic data sets previously generated by forward modelling. Synthetic strain rate distribution was constructed from a linear superposition of analytical functions. *a*, Synthetic subsidence data set where cumulative thickness of strata is sampled every 1 Myr with $\sigma_1 = \pm 25$ m shown by thick line; thin dark line is subsidence curve calculated from initial strain rate distribution. *b*, Initial strain rate distribution determined from gradient of data using equation (2) with $BQ(t)=0$, hence long non-zero tail. *c*, Solution obtained by iterating until thin solid line matches thick line. *d*, Final strain-rate distribution, which agrees with original synthetic distribution to within $<1\%$. *e*, Solution obtained for synthetic subsidence data shown in *a* but cumulative thickness of strata is now sampled every 5 Myr with added random noise (up to ± 100 m of water-loaded subsidence). *f*, Calculated strain rate distributions obtained by inversion of 20 synthetic data sets obtained by adding different random noise to data set in *e*. Note ability to retrieve information about phase and amplitude of strain rate oscillations. *g*, Solution obtained for different synthetic data which only require a single minimum for strain rate distribution. *h*, Calculated strain rate distributions obtained by inversion of 20 synthetic data sets obtained by adding different random noise to the data in *g*.



to determine β , which is related to the vertical strain rate, $G(t)$, by

$$\beta = \exp \left(\int_0^{\Delta t} G(t) dt \right)$$

where Δt is the duration of stretching. If $G(t)$ is constant, $\beta = \exp(G\Delta t)$. The standard expression for theoretical water-loaded subsidence¹² is

$$S(t) = A(1 - 1/\beta) - BQ(t) \quad (1)$$

where

$$A = t_c(\rho_m - \rho_c)/(\rho_a - \rho_w)$$

$$B = \alpha \rho_m / (\rho_a - \rho_w)$$

and

$$Q(t) = \int_0^a [T(z, t) - T(z, \infty)] dz$$

$T(z, t)$ is the temperature of the lithosphere as a function of depth and time and $T(z, \infty)$ is the equilibrium temperature struc-

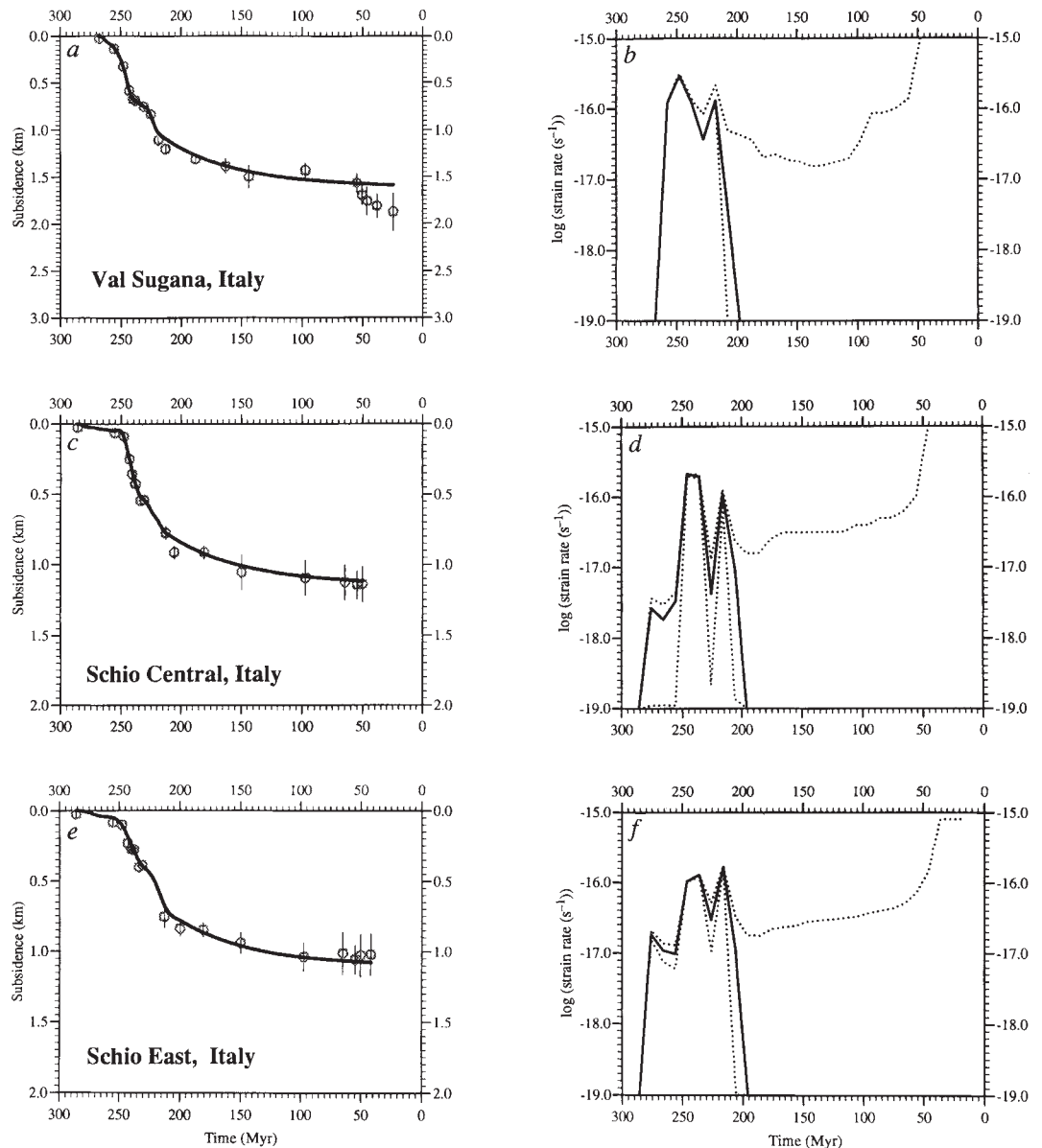
ture of the lithosphere. The symbols and values for other parameters are given in Table 1. $Q(t)$ calculates the difference between the perturbed and equilibrium temperature structure and is necessarily a function of $G(t)$, whereas A and B are constant.

Here, calculating subsidence as a function of time from knowledge of the strain rate variation is defined as the forward problem. The geologically more interesting inverse problem can be solved directly by rearranging equation (1):

$$G(t) = \frac{d}{dt} \left\{ \ln \left(\frac{A}{A - S(t) - BQ(t)} \right) \right\} \quad (2)$$

Given $S(t)$, equation (2) can be solved iteratively to yield $G(t)$ using a starting distribution of $G(t)$ obtained by setting $BQ(t) = 0$ (that is, by ignoring the temperature dependence of density). At each iteration, $Q(t)$ is calculated by standard finite-difference techniques. For synthetic data, convergence is usually rapid and $G(t)$ can be obtained accurately without any *a priori* knowledge of the stretching period. Direct inversion necessitates numerical differentiation and so is susceptible to noise. For discrete and noisy data, the inverse problem is best solved by calculating a range of forward models. $G(t)$ is first parameterized by using M

FIG. 2 Inversion of three stratigraphic sections from the Dolomites, northern Italy (ref. 13). *a, c, e*, Circles, water-loaded subsidence data; vertical bars, 2σ ; thin dark line, best-fitting subsidence curve calculated from final strain rate distribution. *b, d, f*, Thin dark line, linearly interpolated strain rate distribution obtained by inversion and corresponding to best-fitting theoretical subsidence curve; pair of dotted lines, minimum and maximum values of strain rate at any one time that still yield acceptable fit to data (95% confidence level). Where either dotted line is absent, it either coincides with strain-rate curve or with abscissa. Divergence of dotted lines with time indicates decrease in resolution and it is clear that higher strain rates are better resolved. Covariance matrices indicate that strain rates at different times are not significantly correlated. Strain rate minima are well resolved and occur at ~ 230 Myr. Results are not significantly affected either by systematic errors in cumulative thickness of strata (due to stylolitization in carbonates (up to 30%)), or by reducing the thickness of the lithosphere to 100 km. Total strains are as follows; Val Sugana ($\beta = 1.28$), Schio Central (1.25) and Schio East (1.21).



discrete values, G_k , at time intervals δt , where δt is 5–10 Myr. $G(t)$ can then be obtained by interpolation. It is necessary to impose smoothing on $G(t)$ in order to stabilize inversion. Because the problem is non-linear and as all values of G should be >0 , I chose to minimize a trial function by varying G_k . Powell's algorithm¹⁵ was then used to find the minimum of the function H :

$$H = \left[\frac{1}{N} \sum_{i=1}^N \left(\frac{S_i^o - S_i^c}{\sigma_i} \right)^2 \right]^{1/2} + W_1 \left[\frac{1}{M-1} \sum_{k=2}^M \left(\frac{G_k - G_{k-1}}{\delta t} \right)^2 \right]^{1/2} + W_2 \left[\frac{1}{M} \sum_{k=1}^M (G_k'')^2 \right]^{1/2} + \frac{W_3}{M} \sum_{k=1}^M |\log(G_k)| \quad (3)$$

where S_i^o and S_i^c are the observed and calculated subsidences used for inversion, N is the number of observations of subsidence and W_1 , W_2 and W_3 are weighting coefficients. G_k'' are estimates of the second derivative of G generated by cubic spline interpolation. The first term on the right-hand side of equation (3) is zero when calculated and observed values of S_i agree for all of the subsidence data. Dividing the difference between them by σ_i , the variance, causes each term in the summation to have unit variance. The second and third terms cause $G(t)$ to be smooth, and the fourth term tends to infinity smoothly as any G_k approaches zero. All of the results discussed below were obtained using $W_1 = 0.5$, $W_2 = 0.5$ and $W_3 = 0.05$ – 0.5 . Varying the values of W_1 and W_2 by several orders of magnitude has a negligible effect, but W_3 must be decreased during inversion to ensure that strain rates converge to zero where required (for example, during the post-rift period).

The algorithm has been applied to a wide range of synthetic data generated by forward modelling, permitting the resolution of the inverse calculation to be estimated. Some examples are shown in Fig. 1. For perfect data, the calculated variation in $G(t)$ is identical to the original forward modelled input. Progressively discretized data with added random noise have also been inverted and show that stable results are obtained although the phase and amplitude of high-frequency strain rate variations are eventually lost due to aliasing effects. For example, if subsidence data are sampled every 5 Myr, only 1 cycle per 10 Myr of strain rate variation can be retrieved ($\sim 10^{-15}$ Hz). Multiple stretching episodes, each with their own strain rate distribution, can be resolved by multiple step inversion. Error analysis demonstrates that resolution is good during the rifting period, gradually decreasing with time.

The method is also applied to stratigraphic sections from the Dolomites (northern Italy), a region that underwent lithospheric stretching during the Triassic and early Jurassic periods (ref. 13 and references therein). This region has been chosen because platform carbonates predominate during the syn-rift period and so both stratigraphic and palaeobathymetric control are good. Syn-rift subsidence varies over time periods which are significantly longer than the uncertainty in dating. In addition, syn-rift topography is generally modest and so the two-dimensional

effects of a small lithospheric flexural rigidity can be ignored. The subsidence data were first backstripped using the standard method^{16,17}, which assumes Airy isostasy, before inversion (see ref. 13 for details). No *a priori* assumptions concerning rift duration or total strain were applied. As before, the starting solution for $G(t)$ was generated directly from the data using equation (2).

The results for the Val Sugana section are shown in Fig. 2a and b. The duration of rifting is well defined (265–195 Myr) and the total strain, β , is 1.28, in broad agreement with Wooler *et al.*, who used sedimentological and tectonic constraints to delimit the rift period, assuming a constant strain rate. In the first 15 Myr, $G(t)$ climbs from zero to a maximum of $10^{-15.5} \text{ s}^{-1}$. It then decreases to $10^{-16.5} \text{ s}^{-1}$ before rising again to a secondary maximum of 10^{-16} s^{-1} . Termination of stretching is marked by the rapid decrease of strain rate towards zero. Note that the increase in subsidence during the Tertiary era has not been included in the inversion because it is due to Eo-Alpine compression (that is, foreland basin formation)¹³. Sensitivity analysis shows that the strain rate minimum is well resolved, reflecting structure within the observed subsidence data. As expected, resolution decreases with time and so the strain rate distribution during the latter part of the post-rift phase is poorly determined. The difference between a strain rate of 10^{-18} and 0 s^{-1} has a negligible effect on the post-rift subsidence gradient.

The results of inverting two other sections are also shown in Fig. 2. In each case the inverse calculation finds a good fit to the data, regardless of the initial strain rate distribution. The average strain rate is $\sim 10^{-16.5} \text{ s}^{-1}$, in accordance with the results of simple dynamic models³ and seismology⁹. The maximum extensional strain achieved at these low average strain rates are consistent with the predictions of numerical calculations using temperature-dependent rheologies³ and with results determined by moment tensor inversion of earthquake data⁸. The duration of lithospheric stretching, and the timing of the strain rate minimum (~ 230 Myr), are broadly consistent throughout the Dolomites yielding $\beta < 1.3$. Such minima are observed in other sedimentary basins and their existence has been predicted by dynamic models which assume that the lithosphere deforms by power-law creep³. The timing of maximum strain rate generally precedes syn-rift volcanism (Ladinian–Carnian, 235 Myr)¹⁸.

The simple one-dimensional scheme presented here can be generalized and applied to two- and three-dimensional subsidence data from extensional sedimentary basins where the stratigraphy and palaeobathymetry are accurately known. The resultant information concerning spatial and temporal strain rate variation will provide important constraints for dynamic modelling of lithospheric deformation. A complete description of the mechanics of lithospheric deformation requires information about the magnitude of the driving stresses, the rate of deformation and rheology of the lithosphere. It is less likely that a similar approach can be used for foreland basins because of the large degree of uncertainty about load geometry, degree of erosion and flexural rigidity. $\square$

Received 12 January; accepted 4 November 1993.

- England, P. & Jackson, J. A. *Rev. Earth planet. Sci.* **17**, 197–226 (1989).
- Molnar, P. *Nature* **355**, 131–137 (1988).
- Sonder, L. J. & England, P. C. *J. geophys. Res.* **94**, 7603–7619 (1989).
- Le Pichon, X. & Angelier, J. *Tectonophysics* **60**, 1–42 (1979).
- Heard, H. C. *Phil. Trans. R. Soc. A* **283**, 173–186 (1976).
- England, P. C. *J. geophys. Res.* **88**, 1145–1152 (1983).
- Pfiffner, O. A. & Ramsay, J. G. *J. geophys. Res.* **87**, 311–321 (1982).
- Jackson, J. & McKenzie, D. *Geophys. J.* **93**, 45–73 (1988).
- Holt, W. E., Ni, J. F., Wallace, T. C. & Haines, A. J. *J. geophys. Res.* **96**, 14595–14632 (1991).
- Billiris, H. *et al. Nature* **350**, 124–129 (1991).
- McKenzie, D. *Earth planet. Sci. Lett.* **40**, 25–32 (1978).
- Jarvis, G. T. & McKenzie, D. *Earth planet. Sci. Lett.* **48**, 42–52 (1980).

- Wooler, D. A., Smith, A. G. & White, N. J. *J. geol. Soc. Lond.* **149**, 517–532 (1992).
- Barton, P. & Wood, R. *Geophys. J. R. astr. Soc.* **79**, 987–1022 (1984).
- Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. *Numerical Recipes* (Cambridge Univ. Press, 1986).
- Steckler, M. S. & Watts, A. B. *Earth planet. Sci. Lett.* **41**, 1–13 (1978).
- Slater, J. G. & Christie, P. A. F. *J. geophys. Res.* **85**, 3711–3739 (1980).
- Sloman, L. E. *J. geophys. Res.* **94**, 4655–4666 (1989).
- Cochran, J. R. *Earth planet. Sci. Lett.* **46**, 266–274 (1980).

ACKNOWLEDGEMENTS. I thank H. Kerr, D. Latin, L. Lonergan, D. McKenzie, M. Sambridge and S. Singh for discussions, D. Wooler for permitting the use of his subsidence data and P. England for providing a helpful review. The support of the Department of Earth Sciences at Cambridge University is acknowledged.

Electrical conductivity of the Earth's lower mantle

T. J. Shankland^{*†}, J. Peyronneau^{*} & J.-P. Poirier^{*}

^{*} Département des Géomatériaux, Institut de Physique du Globe de Paris, 4 place Jussieu, 75252 Paris Cedex 05, France

THE electrical conductivity of the Earth's lower mantle constrains both the propagation to the surface of geomagnetic disturbances in the core and the nature of core-mantle coupling. Extrapolations of laboratory measurements on materials representative of the lower mantle agree weakly^{1,2} or not at all^{3,4} with recent geophysical models⁵⁻⁸ of lower-mantle electrical conductivity based on variations of magnetic and electrical fields measured at the Earth's surface. Here we report d.c. conductivity measurements on samples with compositions approximating that of the lower mantle, at pressures of 1.2 to 40 GPa and temperatures in the range 20 to 400 °C. Our results agree with some of those obtained previously^{1,2}. But in contrast to this previous work, we extrapolate the results to lower-mantle conditions by adopting a functional form for the conductivity that incorporates the effect of pressure as well as temperature. The resulting estimates of conductivity are in agreement with the geophysical determinations⁵⁻⁸. We find that, because of a very weak dependence on temperature, pressure and composition, the conductivity is likely to vary by no more than about a factor of five across the entire lower mantle, reaching a maximum value of only 3–10 S m⁻¹. Lateral temperature variations as large as a few hundred degrees will therefore be hard to detect geophysically, and the compositionally distinct D'' layer at the base of the lower mantle remains the only possible location for a highly conducting layer.

We used three starting materials from San Carlos, Arizona to represent possible lower-mantle compositions. The samples were an enstatite (Mg_{1-x}, Fe_x)SiO₃ with an iron content of 11 atom% ($x=0.11$) and two olivines (Mg_{1-x}, Fe_x)₂SiO₄ with $x=0.11$ and 0.16, respectively. Samples were ground under alcohol in an agate mortar to a grain size of <5 µm. We used either freshly ground samples dried at 120 °C or previously ground powders heated to above 300 °C, procedures found necessary to remove the influence of adsorbed water². The powder was then loaded between two tungsten electrodes in a diamond-anvil cell. An external furnace was used to heat the cell, and temperature was monitored by a thermocouple in contact with one of the diamonds^{1,2}. Each sample was taken to a pressure of ~40 GPa, and a strip of material connecting the electrodes was transformed by laser heating into its high-pressure phase (enstatite adopting a perovskite structure and olivine a mixed phase of perovskite and magnesio-wüstite). Conductivities were measured as a function of temperature from 20 to 400 °C at constant pressures of 1.2–40 GPa.

Lower-mantle pressures increase from 23 GPa at 670 km depth to 136 GPa (ref. 9) at 2,885 km, and the corresponding temperature range (neglecting a probable thermal boundary layer at the base of the mantle) is ~1,900 K to perhaps 2,800–3,400 K (ref. 10). Given experimental limitations on temperatures and pressures we expressed electrical conductivity σ in a manner that allows extrapolation and interpolation of laboratory data to mantle conditions. A functional form to accommodate effects of pressure and temperature is the Arrhenius equation for semiconductors with an activation enthalpy ΔH replacing activation energy ΔU ,

$$\sigma = \sigma_0 e^{-\Delta H/kT} \quad (1)$$

where $\Delta H = \Delta U + P\Delta V$, k is Boltzmann's constant, ΔV is activation volume (which expresses the effect of pressure on the energy barrier to charge transport) and σ_0 is the limiting value of con-

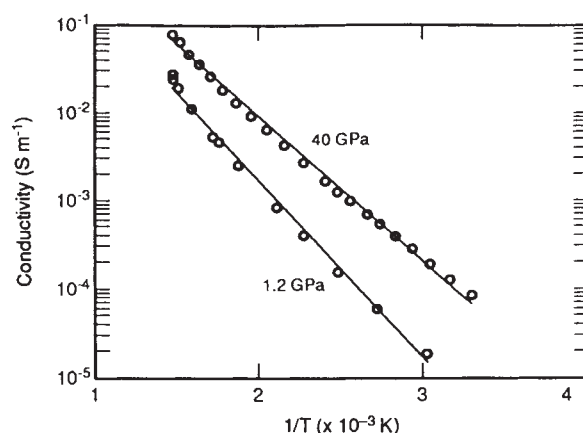


FIG. 1 Electrical conductivity (σ) plotted against reciprocal temperature of transformed sample having olivine bulk composition and $x=0.16$. Solid line shows best simultaneous three parameter fit to $\ln \sigma$ for runs at 1.2 and 40 GPa.

ductivity as temperature T approaches infinity or T^{-1} approaches zero. The quantity σ_0 incorporates the temperature-independent entropy term in the activation Gibbs free energy but may also depend on pressure P , iron content x and oxygen activity a_{O_2} (refs 1, 3, 11). Activation volume ΔV is calculated from the change in ΔH with pressure (by determining the slope of $\ln(\sigma)$ against T^{-1} for more than one pressure)

$$\Delta V = \left(\frac{\partial \Delta H}{\partial P} \right) = - \frac{\partial \left(\frac{\partial \ln \sigma}{\partial T^{-1}} \right)_P}{\partial P} \quad (2)$$

Figure 1 shows results at pressures of 1.2 and 40 GPa for magnesio-wüstite-perovskite with $x=0.16$. We determined σ_0 , ΔU and ΔV by simultaneously fitting conductivity data from the two pressure runs to equation (1) with a nonlinear Levenberg-Marquardt method¹². This procedure gives the same result for equation (2) if σ_0 is pressure-independent. Separate straight-line fits to the two data sets of Fig. 1 have intercepts $\sigma_0=21$ and 17 S m⁻¹, a variation within experimental error. In this case we could treat σ_0 as a constant, which is not possible in general because it may vary substantially^{3,4,13}. Table 1 summarizes results for the three materials studied.

Properties of the three materials have several features in common: they all have σ_0 of the order of 10 S m⁻¹, ΔU of ~0.4 eV and activation volumes that are both small (0.1–0.2 cm³ mol⁻¹) and negative. ($\Delta V = -0.1$ cm³ mol⁻¹ of magnesio-wüstite-perovskite with $x=0.11$ agrees with the value² for a similar sample calculated from $(\partial \ln \sigma / \partial P)_T = -\Delta V/kT$, which follows from equation (1) only for σ_0 independent of P .) Negative activation volumes of this magnitude are consistent with a conduction mechanism in which charge is carried by electrons hopping from Fe²⁺ to Fe³⁺ ions (the small polaron model^{1,2,14}), although it is plausible that any electronic mechanism could have a ΔV of this order. If conduction were ionic¹⁵, a larger, positive ΔV would be expected because ionic motion requires lattice dilation as is

TABLE 1 Fitting parameters to equation (1)

Bulk composition of sample	$\ln \sigma_0$	σ_0 (S m ⁻¹)	ΔU (eV)	ΔV (cm ³ mol ⁻¹)
(Mg _{0.84} , Fe _{0.16}) ₂ SiO ₄	2.90 ± 0.24	18	0.40 ± 0.01	-0.18 ± 0.01
(Mg _{0.89} , Fe _{0.11}) ₂ SiO ₄	2.82 ± 0.56	17	0.42 ± 0.02	-0.10 ± 0.02
(Mg _{0.89} , Fe _{0.11})SiO ₃	1.92 ± 0.68	7	0.48 ± 0.02	-0.26 ± 0.03

Indicated errors are 95% confidence limits ($\pm 2 \times$ standard deviation) with no error assignment to the data. σ is conductivity, σ_0 is the limiting value of conductivity as temperature T approaches infinity (or T^{-1} approaches zero), P is the pressure, ΔV is the activation volume, ΔU is the activation energy and k is Boltzmann's constant.

[†] Present address: Earth and Environmental Sciences Division, MS D443, Los Alamos National Laboratory, Los Alamos, New Mexico 84545, USA.

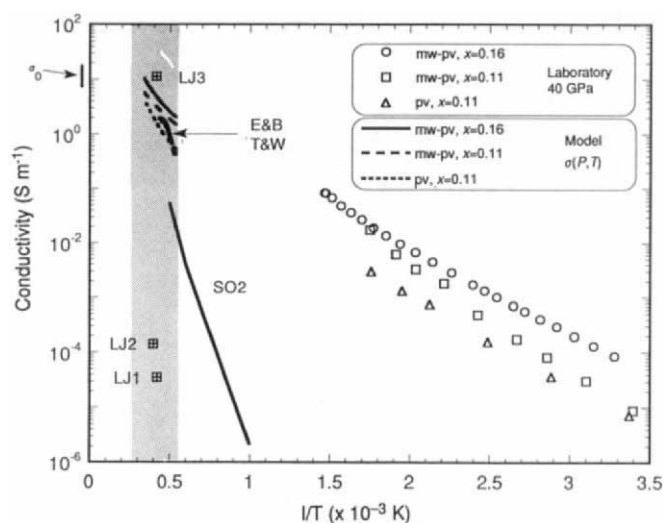


FIG. 2 Electrical conductivity (σ) plotted against reciprocal temperature for three laboratory samples at 40 GPa (hollow symbols) and for lower-mantle conductivity models based on laboratory data (lines); pv is perovskite and mw-pv is magnesiowüstite-perovskite. The two sets of geophysical models shown as E&B (ref. 5) and T&W (ref. 6) overlap. For comparison, data of Li & Jeanloz^{3,4} are indicated: LJ1 is magnesiowüstite-perovskite, $x=0.12$; LJ2 is perovskite, $x=0.12$; LJ3 is magnesiowüstite-perovskite, $x=0.20$. Standard olivine SO2 (ref. 20) at $P=0$ illustrates the far greater effect of T when ΔH is large (1.5–4 eV). High temperature points are deliberately left bunched to indicate how small are their differences at mantle temperatures on a T^{-1} scale. The hatched area indicates temperatures in the range from 1,900 K to 4,000 K.

the case for diffusion¹⁶. In another extrinsic mechanism Fe^{2+} could act as a donor for conduction-band electrons, or Fe^{3+} as an acceptor to produce valence-band holes (a variation on a suggested model³). For olivine that has disproportionated into perovskite ($x=0.06$) and magnesiowüstite ($x=0.16$) (ref. 17), a third model is conduction through the iron-rich interconnected magnesiowüstite^{18,19}. Of the two samples having the same bulk iron content ($x=0.11$), the magnesiowüstite-perovskite mixture has slightly higher conductivity than the perovskite, and this mechanism probably accounts for the difference.

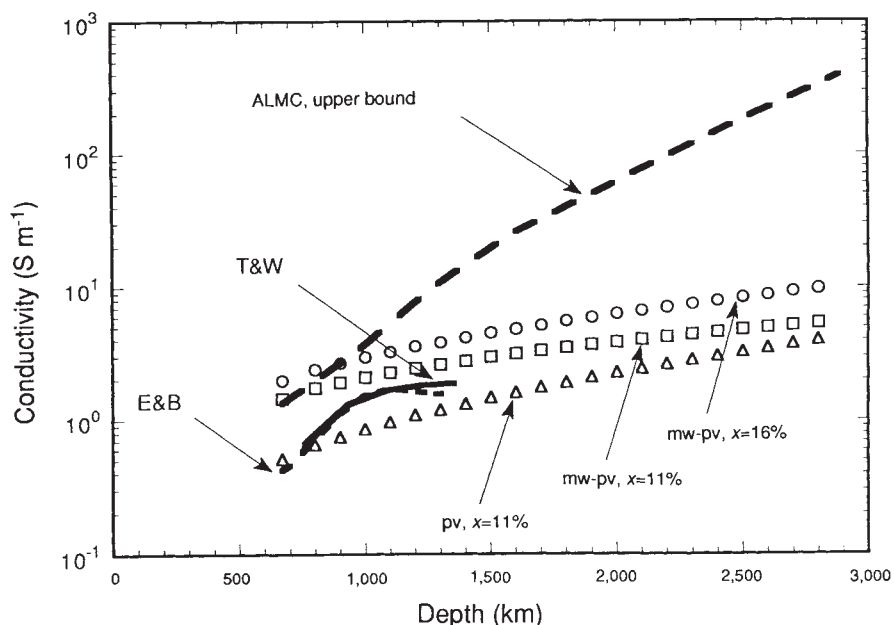
For geophysical purposes, it is less important to determine the conduction mechanism than to demonstrate the similarities

between conductivities based on these laboratory measurements and mantle conductivities based on geomagnetic data. At the high temperatures and pressures of the lower mantle the exponent in equation (1) is small, and the effects of both P and T on σ are weak. Further, equation (1) shows the effect of pressure on σ is smaller at mantle temperatures than at room temperature. Conductivity should therefore approach σ_0 , and the parameters that govern σ_0 (such as x and a_{O_2}) should become at least as important as T and P in the lower mantle.

Using the results in Table 1 in equation (1) we have extrapolated the conductivities of the three materials to pressures and temperatures appropriate for the lower mantle. Pressures as a function of depth for our models are taken from the Earth model PREM (ref. 9) and temperatures from Shankland and Brown¹⁰. Despite the large temperature span between laboratory and mantle conditions, extrapolation of the experimental data on a T^{-1} scale is not great. Figure 2 shows the resulting models, along with our 40 GPa experimental curves, a zero-pressure conductivity model for isotropic olivine (SO2) (ref. 20) and conductivity data from Li and Jeanloz^{3,4} for lower-mantle phases having high and low iron contents. Also shown are two geophysical conductivity models^{5,6}. Two points are immediately clear from Fig. 2: first, both geophysical and experimental models give similar values, probably within their errors of calculation, for lower-mantle conductivity. Second, the experimental models are close to their limiting conductivity σ_0 as expected. Even though temperature increases by $\sim 50\%$ and pressure by a factor of ~ 6 across the lower mantle, there is simple little latitude for their variations as functions of depth to have much effect on σ . The clustering seen in these models results from similar values of σ_0 and is relatively insensitive to ΔU and ΔV . Choosing a temperature profile higher by a few hundred degrees (K) would not appreciably affect the result.

Figure 3 shows an expanded version of the geophysical conductivity models and our laboratory models (the upper left corner of Fig. 2) plotted on a scale of depth instead of T^{-1} . In the outer lower mantle two overlapping magnetotelluric (MT) models from the southwestern USA³ and from global data⁶ are close to an upper bound model²¹ based on combined geomagnetic deep sounding (GDS) and the frequency spectrum of core geomagnetic variations. Not shown are a radially symmetric GDS model⁷ and a MT/GDS model⁸ beneath the Canadian shield that match the first two curves. As can be seen, the laboratory-based models for perovskite and for mixed magnesiowüstite-perovskite compositions having $x=0.11$ are

FIG. 3 Electrical conductivity (σ) plotted against depth in the lower mantle for extrapolated laboratory models (hollow symbols); pv is perovskite and mw-pv is magnesiowüstite-perovskite. Geophysical models are shown as E&B (ref. 5) (dotted line), T&W (ref. 6) (solid line), and the upper-bound curve ALMC (ref. 21) (dashed line).



close to each other and to the geophysical models. The accuracies of the independent laboratory and field models do not permit forcing a given mantle composition, but Fig. 3 does show that there is adequate quantitative agreement between them, and that this will still be the case if we allow for minor variations in the parameters in Table 1. The disagreement shown in Fig. 2 between these results and the low conductivities of Li and Jeanloz⁴ persists for samples of low iron content (although agreement is better for higher x), and this probably reflects the different experimental methods used. But it is clear from Fig. 3 that lower-mantle conductivity inferred from field measurements more closely resembles that of materials such as those in Table 1.

These results permit the following conclusions: (1) a magnesiowüstite-perovskite mixture having an iron content of ~ 10 atom% gives a good representation of electrical conductivity down to the base of the mantle, with the exception of the bottom 100–200 km (the D'' layer), which is believed from seismic models²² to be of a heterogeneous composition. (2) Lateral conductivity variations should be extremely hard to resolve geophysically, a prediction consistent with similarities of recent geophysical models^{5–8}. Over the entire lower mantle, conductivity increases by only a factor of five so that lateral temperature variations of ± 200 K are not likely to produce resolvable departures from radial symmetry unless they are also accompanied by strong compositional heterogeneity, for instance, by volatiles in plumes derived from the D'' layer. Conversely, if lateral variations can be established, they may be good evidence for such heterogeneity. (3) Electromagnetic induction of currents in the mantle by magnetic variations from the core could exert torques on the mantle with possible effects on Earth's orbital parameters, and the spectrum of core magnetic fluctuation is filtered by propagation through a conducting medium. Should higher conductivity in the lower mantle be needed for core-mantle coupling or for restricting the frequency range of electromagnetic disturbances propagated outward from the core, then the D'' layer may be the only other place in which to locate a more conducting zone because of its different physical properties. Conversely, these low conductivities combined with a poorly conducting model²³ for the D'' layer would suggest negligible electromagnetic core-mantle coupling and that the observed frequency spectrum of core disturbance is an adequate representation of the timescale of core motions. $\square$

A siderophore from a marine bacterium with an exceptional ferric ion affinity constant

Richard T. Reid*, David H. Live†, D. John Faulkner‡ & Alison Butler*§

* Department of Chemistry, University of California, Santa Barbara, California 93106-6090, USA

† Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, New York 10021, USA

‡ Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093-0212, USA

§ To whom correspondence should be addressed.

VIRTUALLY all microorganisms require iron for growth. The paucity of iron in surface ocean water (~ 0.02 – 1.0 nM (refs 1, 2)) has spurred a lively debate concerning iron limitation of primary productivity^{3–6}, yet little is known about the molecular mechanisms used by marine microorganisms to sequester iron. Terrestrial bacteria use a siderophore-mediated ferric uptake system⁷. A siderophore is a low-molecular-mass compound with a high affinity for ferric ion which is secreted by microorganisms in response to low-iron environments; siderophore biosynthesis is regulated by iron levels, with repression by high iron. Although open-ocean marine microorganisms (such as phytoplankton⁸ and bacteria⁹) produce siderophores, the nature of these siderophores has not been investigated. We report here the first structure determination, to our knowledge, of the siderophores from an open-ocean bacterium, alterobactin A and B from *Alteromonas luteoviolacea*. *A. luteoviolacea* is found in oligotrophic¹⁰ and coastal¹¹ waters. Alterobactin A has an exceptionally high affinity constant for ferric ion. We suggest that at least some marine microorganisms may have developed higher-affinity iron chelators as part of an efficient iron-uptake mechanism which is more effective than that of their terrestrial counterparts.

Alterobactin A (1) and alterobactin B (2) were isolated and purified as previously described¹². The molecular formulae of

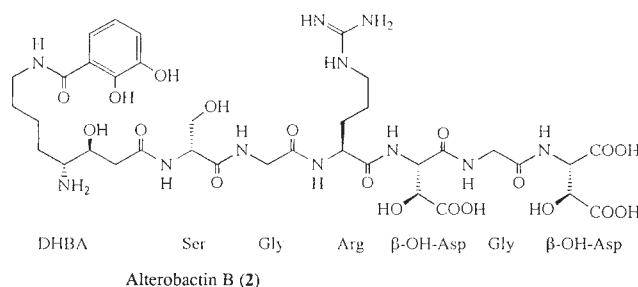
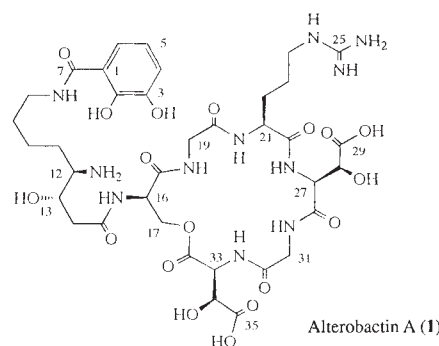


FIG. 1 Structures of alterobactin A (1) and alterobactin B (2). The absolute configurations of D-serine, L-arginine, L-threo-β-hydroxyaspartic acid and 4S-amino, 3R-hydroxy moiety of 4,8-diamino-3-hydroxyoctanoic acid in (1) and (2) were established by standard procedures^{16,24,25}.

Received 10 May; accepted 13 October 1993.

- Peyronneau, J. & Poirier, J.-P. *Nature* **342**, 537–539 (1989).
- Poirier, J.-P. & Peyronneau, J. in *High-Pressure Research: Application to Earth and Planetary Sciences* Geophys. monogr. No. 67 (eds Syono, Y. & Manghnani, M. H. 77–87 (Am. Geophys. Un., Washington DC, 1992).
- Li, X. & Jeanloz, R. *J. geophys. Res.* **96**, 6113–6120 (1991).
- Li, X. & Jeanloz, R. *J. geophys. Res.* **95**, 5067–5078 (1990).
- Egbert, G. D. & Booker, J. R. *J. geophys. Res.* **97**, 15099–15112 (1992).
- Tarits, P. & Wahr, J. *EOS* **73**, 523 (1992).
- Constable, S. J. *Geomag. Geoelectr.* **45**, 1–22 (1993).
- Schultz, A., Kurtz, R. D., Chave, A. D. & Jones, A. G. *Geophys. Res. Lett.* (in the press).
- Dziewonski, A. M. & Anderson, D. L. *Phys. Earth planet. Inter.* **25**, 297–356 (1981).
- Shankland, T. J. & Brown, J. M. *Phys. Earth planet. Inter.* **38**, 51–58 (1985).
- Hirsch, L. M. & Shankland, T. J. *Geophys. Res. Lett.* **18**, 1305–1308 (1993).
- Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. *Numerical Recipes* 498–538 (Cambridge Univ. Press, 1986).
- Li, X., Ming, L.-C., Manghnani, M. H., Wang, Y. & Jeanloz, R. *J. geophys. Res.* **98**, 501–508 (1993).
- Hirsch, L. M. & Shankland, T. J. *Geophys. Res. Lett.* **18**, 1305–1308 (1991).
- Gautason, B. & Muehlenbachs, K. *Science* **260**, 518–521 (1993).
- Misener, D. J. in *Geochemical Transport and Kinetics* (eds Hofmann, A. W., Giletti, B. J., Yoder, H. S. Jr. & Yund, R. A.) 117–129 (Carnegie Instn of Wash., Washington DC, 1974).
- Guyot, F., Madon, M., Poirier, J.-P. & Peyronneau, J. *Earth planet. Sci. Lett.* **90**, 52–64 (1988).
- Wood, B. J. & Nell, J. *Nature* **351**, 309–311 (1991).
- Li, X. & Jeanloz, R. *J. geophys. Res.* **95**, 21609–21612 (1990).
- Constable, S., Shankland, T. J. & Duba, A. G. *J. geophys. Res.* **97**, 3397–3404 (1992).
- Achache, J., Le Mouél, J. L. & Courtillot, V. *Geophys. J. R. ast. Soc.* **65**, 579–601 (1981).
- Lay, T. *EOS* **70**, 49–59 (1989).
- Poirier, J.-P. & le Mouél, J. L. *Phys. Earth planet. Inter.* **73**, 29–37 (1992).

ACKNOWLEDGEMENTS. We thank P. Tarits, A. Schultz, R. Kurtz, A. Chave and A. Jones for making data available in advance of publication. This work was supported by the US Department of Energy, the French programme Dynamique et Bilan de la Terre and CNRS. T.J.S. thanks the CNRS and the Institut de Physique due Globe de Paris for additional support.

(1) $C_{36}H_{53}N_{11}O_{18}$ and (2) $C_{36}H_{55}N_{11}O_{19}$ were established from ^{13}C NMR data and high-resolution fast-atom bombardment (FAB) mass spectral data: $(M+H)^+$ peak at $m/z = 928.3954$ for (1) and $(M+H)^+$ peak at $m/z = 946.3961$ (and its deuterium-exchanged derivative, $(M+H)^+ = 968$ and $(M+D)^+ = 969$) for (2). The 1H and ^{13}C NMR data were assigned as shown in Table 1. All the amide linkages were defined by the coupling of the carbonyl ^{13}C resonance to the 1H of $C\alpha$, and in some cases to $C\beta$ (ref. 13 and Table 1); the position of dihydroxybenzoate was established by 1H - ^{13}C coupling of H6-C7 and H8-C7 (ref. 13 and Table 2). β -Hydroxyaspartate was identified by the ^{13}C

chemical shifts of the β -carbons and the proton-splitting pattern, consistent with values in refs 14, 15 and the NMR spectrum of an authentic sample. The 4,8-diamino-3-hydroxyoctanoic acid derivative was identified by vicinal 1H - 1H couplings (13H-14H, 13H-12H, 12H-11H, 11H-10H, 10H-9H and 9H-8H) and long-range ^{13}C - 1H couplings (7C-8H, 9C-8H, 9C-10H, (10,11)C-(9,10,11)H, 11C-12H, 11C-13H, 15C-14H), in agreement with ref. 16. The lactam in (1) was identified by: 1) the superCOSY spectrum¹⁷ showing coupling between the serine β -proton (H17) and α -proton (H33) of the terminal β -hydroxyaspartic acid; 2) reduction by 18 mass units and one acid titratable proton rela-

TABLE 1 1H and ^{13}C resonances for alterobactin A and B

(1) (alterobactin A)			(2) (alterobactin B)		
	^{13}C (δ)	1H (δ)		^{13}C (δ)	1H (δ)
					^{13}C - 1H Long-range couplings
Dihydroxybenzoic acid					
C1	116.67	—	116.95	—	H5
C2	146.99	—	146.95	—	H4, H6
C3	144.55	—	144.61	—	H5
C4	119.56	7.04	119.37	7.07	H6
C5	119.24	6.82	119.70	6.86	
C6	118.68	7.21	118.84	7.24	H4
C7	170.09	—	170.17	—	H6, H8
4,8-Diamino-3-hydroxyoctanoic acid					
C8	38.85	3.39	38.99	3.36	
C9	28.08	1.7	28.15	1.7	H8, H10
C10	22.24	1.7	22.34	1.7	H8, H9, H11, H12
C11	26.34	1.7	26.47	1.7	H12, H(10)
C12	55.20	3.39	55.23	3.36	H11, H14
C13	67.20	4.36	67.34	4.36	H11, H12, H14
C14	37.82	2.59	37.90	2.56	
C15	172.54	—	172.55	—	H13, H14, H16
Serine					
C16	51.69	4.78	55.87	4.46	H17
C17	64.80	4.41, 4.66	61.08	3.88	H16
C18	170.90	—	172.32	—	H16, H17, H19
Glycine-1					
C19	42.46	4.04	42.49	3.98	
C20	172.37	—	171.58	—	H19, H21
Arginine					
C21	55.52	4.13	53.61	4.36	H22
C22	27.46	1.7	27.94	1.7	H21, H23, H24
C23	24.05	1.7	24.28	1.7	H21, H22, H24
C24	40.41	3.20	40.51	3.13	H22, H23
C25	156.70	—	156.72	—	H24
C26	174.73	—	173.08	—	H21, H27
β -Hydroxyaspartate-1					
C27	56.29	5.03	55.73	4.99	
C28	70.20	4.93	70.25	4.82	
C29	174.61	—	174.26	—	H27, H28
C30	171.58	—	171.05	—	H27, H28, H31
Glycine-2					
C31	42.26	4.08	42.44	4.04	
C32	170.99	—	170.67	—	H31, H33
β -Hydroxyaspartate-2					
C33	55.40	5.16	55.32	5.06	
C34	70.44	4.83	70.40	4.86	
C35	174.24	—	174.41	—	H33, H34
C36	169.48	—	173.85	—	H33, H34

1H and ^{13}C resonances for alterobactin A and B in D_2O . The 1H and ^{13}C NMR data were assigned as shown, by interpretation of the 1H - 1H correlated spectroscopy (COSY), chemical shift correlation by magnetization transfer (CSCM; $J = 145$ Hz), heteronuclear multiple quantum coherence (HMQC; $J = 150$ Hz) and heteronuclear multiple bond correlations (HMBC; $J = 4$ Hz) results. 1H - ^{13}C connectivities were determined using the HMBC pulse sequence of ref. 13. Chemical shift assignments are relative to internal sodium trimethyltetrauteropropionate (TSP) for proton and relative to external $CDCl_3$ for carbon resonances. NMR spectra were run on General Electric GN 500 MHz and Varian Unity 500 MHz instruments. Sample concentrations ranged from 20–50 ml^{-1} . All 1H resonances listed at 1.7 p.p.m. are part of a large cluster between 1.5–1.9 p.p.m.

tive to (2); 3) shifts in the serine α - and β -proton resonances; and 4) the downfield shift of the serine β -carbon (61.08 p.p.m.) in (2).

Iron(III) coordination behaviour distinguishes (1) from (2). Job's method of continuous variations¹⁸ shows a maximum absorbance intensity at 480 nm at a 1/1 ratio of Fe^{3+} /1, but at a 1/2 ratio of Fe^{3+} /2, which establishes that Fe^{3+} is coordinated by one equivalent of (1) and two equivalents of (2). Although both (1) and (2) satisfy the conventional definition of a siderophore, (2) may not be a true siderophore because it forms by base hydrolysis of (1) in the absence of bound iron(III). Iron coordination by (1) occurs by catecholate and both β -hydroxyaspartate moieties. Structure predictions based on molecular modelling calculations show that the iron-binding ligands are precisely poised to coordinate iron, with very little ligand rearrangement on iron binding; any change of absolute configuration of any centre is detrimental to iron binding (Fig. 2).

The striking feature of (1) is the extraordinary affinity for ferric ion. The values of the proton-dependent stability constant, $K_{\text{Fe(1)}}^*$ (ref. 19), determined from reaction of $\text{Fe(III)}-(1)$ with EDTA or $\text{Fe(III)}-\text{EDTA}$ with (1), agree within experimental error, $10^{20.7(0.1)}$ and $10^{20.4(0.2)}$, respectively, indicating that equilibration can be approached from both directions (Table 2). The proton-independent ferric ion stability constant calculated from $K_{\text{Fe(1)}}^*$ and the estimated K_a values of the coordinating ligands (see Table 2 caption) is $10^{49}-10^{53}$, which is not exceeded by any other known siderophore, with the possible exception of enterobactin (10^{49} ; ref. 20) and is more than 15 orders of magni-

tude above any of the other siderophores (Table 2). The seawater pM value ($-\log [\text{Fe}^{3+}]^{21}$ at pH 8.3, $1 \mu\text{M} [\text{Fe}^{3+}]_{\text{tot}}$ and $10 \mu\text{M} [\text{H}_6\text{L}]_{\text{tot}}$ is 30.4, which shows that very little iron(III) would be uncomplexed under these conditions.

Alterobactin A (1) has an extraordinary affinity for iron(III). It is likely that the siderophores produced by open-ocean marine bacteria that live in an extremely low-iron environment may well have evolved unique structures with exceptional iron(III) affinities. Marine phytoplankton, which produce siderophores⁸, may have also developed higher-affinity iron chelators as part of their iron-uptake mechanisms. The structures of only two other marine siderophores have been determined, both of which come from bacteria that live in relatively iron-rich environments: anguibactin from the fish gut bacterium *Vibrio anguillarum*²² and bisucaberin from the deep-sea mud bacterium *Alteromonas haloplanktis*²³. Although the ferric affinity constants of these siderophores have not been reported, they may not be unusual given the surfeit of iron. High ferric affinity constants for sidero-

TABLE 2 Determination of the proton-independent ferric binding constant and pM of alterobactin A (1) and comparison to other siderophores

Alterobactin A (1)					
EDTA:(1)	Abs. λ 476 nm (Std Dev)	[Fe-(1)]	$K_{\text{Fe(1)}}^*$	$K_{\text{Fe(1)}}^{\text{PI}}$ (proton independent)	Ref.
Fe-(1) + EDTA					
1:2	0.154 (0.002)	73.3 μM	$10^{20.7(0.1)}$	$10^{49}\text{--}10^{53}$	This work
1:1	0.119 (0.008)	56.4 μM			
2:1	0.079 (0.005)	38.0 μM			
5:1	0.049 (0.006)	23.0 μM			
10:1	0.0265 (0.0007)	12.5 μM			
FeEDTA + (1)					
1:1	0.096 (0.009)	46.0 μM	$10^{20.4(0.2)}$		
2:1	0.064 (0.010)	30.4 μM			
5:1	0.035 (0.008)	17.2 μM			
10:1	0.0175 (0.0007)	8.5 μM			
Other siderophores					
Enterobactin				10^{49}	20
Ferrioxamine B				$10^{30.99}$	26
Pyoverdine PaA				$10^{30.8}$	27
Ferrichrome				$10^{29.1}$	28, 29
Aerobactin				$10^{22.5}$	30

Solutions of $\text{Fe(III)}-(1)$ with EDTA or $\text{Fe(III)}-\text{EDTA}$ with (1) were allowed to react for 20 h under conditions of 0.1 mM Fe(III) and 0.1 mM (1), 0.5 to 10 equivalents of EDTA in 0.1 M buffer, pH 6.0, using triply distilled H_2O . The absorbance values at a particular EDTA/(1) ratio are averages of eight separate experiments, except for 1:2 and 10:1, which are an average of six and five experiments, respectively. K_{measured} was calculated using the concentration of $\text{Fe(III)}-(1)$ determined spectrophotometrically: λ_{max} 476 nm, ϵ 2,050 $\text{M}^{-1} \text{cm}^{-1}$. $K_{\text{Fe(1)}}^*$, the proton-dependent stability constant, was calculated from the expression for K_{measured} using a value for K_{EDTA}^* of $10^{20.47}$. The proton-independent stability constant ($K_{\text{Fe(1)}}^{\text{PI}}$) for Fe(1) was then calculated from the expressions for $K_{\text{Fe(1)}}^*$ (using an average of the experimental numbers) and $\alpha_{(1)}^{6-}$, assuming the acid dissociation constants for the *ortho*- and *meta*-hydroxyl protons of catechol are 8.4 and 12.1, (K_1, K_2)³¹, respectively, and the $\text{p}K_a$ for the carboxylate proton of each β -hydroxyaspartic acid is 3.5 (K_3, K_4). The $\text{p}K_a$ of the hydroxyl proton of β -hydroxyaspartic acid, however, is not known, but a reasonable estimate would be that of an alcohol, that is, $\text{p}K_a$ 16–18 or higher (K_5, K_6). Thus the lower limit of ($K_{\text{Fe(1)}}^{\text{PI}}$) is 10^{49} (using a $\text{p}K_a$ of 16) to 10^{53} (using a $\text{p}K_a$ of 18).

$K_{\text{measured}} = \frac{[\text{Fe(1)}][\text{EDTA}]}{[\text{Fe(EDTA)}][\text{(1)}]} = \frac{K_{\text{Fe(1)}}^*}{K_{\text{FeEDTA}}^*}$ EDTA' and 1' are the total concentrations of the ligand in all protonation states.

$$K_{\text{FeEDTA}}^* = K_{\text{EDTA}}^* \alpha_{\text{EDTA}}^{4-}$$

K_{EDTA}^* is the proton independent binding constant for $\text{Fe(EDTA)} = 10^{25.1}$ (ref. 32). $\alpha_{\text{EDTA}}^{4-}$ is the fraction of fully deprotonated EDTA in solution.

$$K_{\text{Fe(1)}}^* = K_{\text{Fe(1)}}^{\text{PI}} \alpha_{(1)}^{6-}$$

$\alpha_{(1)}^{6-}$ is the fraction of fully protonate (1) in solution.

The general form for the fraction of deprotonated ligand in solution is

$$\alpha_{\text{H}_n\text{ligand}}^{n-} = \frac{K_1 K_2 \dots K_n [\text{H}^+]^{n-j}}{[\text{H}^+]^n + K_1 [\text{H}^+]^{n-1} + K_1 K_2 [\text{H}^+]^{n-2} + \dots + K_1 K_2 \dots K_n}$$

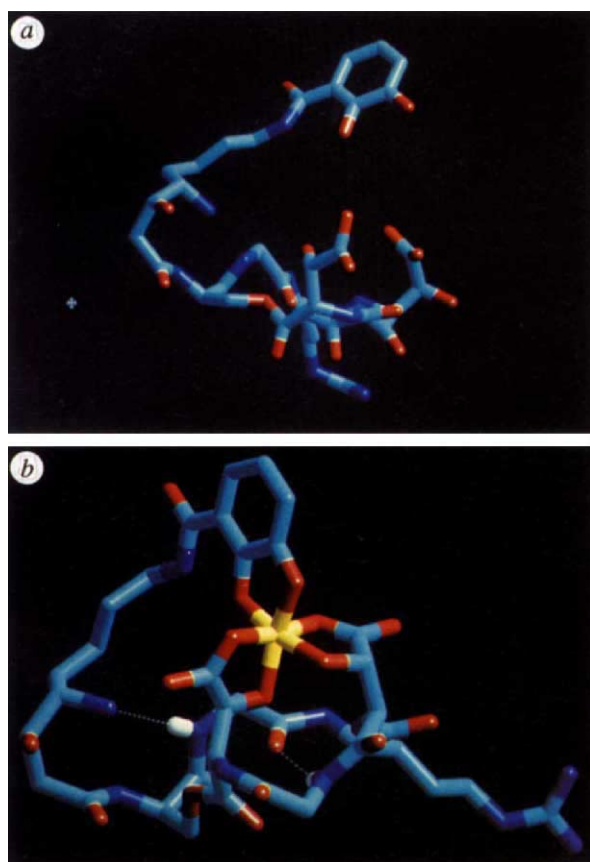


FIG. 2 Predicted conformation of alterobactin A and the iron(III)-alterobactin A complex using HyperChem (Autodesk, Inc). The predicted conformation of alterobactin A (a) was obtained by minimization of the iron-free form using MM+ on a Silicon Graphics Iris 4D/240 GTX workstation. Iron(III) was then added and re-minimized to give structure b. a Shows that the catechol and both β -hydroxyaspartic acid residues create an Fe(III) binding site; 12-amino,13-hydroxy coordination was ruled out by amine modification experiments.

phores from oceanic microorganisms may provide a competitive advantage over other microorganisms with less efficient mechanisms for iron acquisition. We are continuing our investigations into whether the other aspects of the iron-uptake machinery in marine microorganisms (such as Fe(III)-siderophore binding and recognition of outer membrane receptor proteins, iron transport and iron regulation) also show adaptations that favour iron acquisition in extremely low-iron environments. □

Received 10 March; accepted 1 September 1993.

- Bruland, K. W., Donat, J. R. & Hutchins, D. A. *Limnol. Oceanogr.* **36**, 1555–1577 (1991).
- Martin, J. H., Gordon, R. M. & Fitzwater, S. E. *Limnol. Oceanogr.* **36**, 1793–1802 (1991).
- Martin, J. H. & Fitzwater, S. E. *Nature* **331**, 341–343 (1988).
- Price, N. M., Andersen, L. F. & Morel, F. M. M. *Deep Sea Res.* **38**, 1361–1378 (1991).
- Martin, J. H., Gordon, R. M. & Fitzwater, S. E. *Nature* **345**, 156–158 (1990).
- Sunda, W. G., Swift, D. G. & Huntsman, S. A. *Nature* **351**, 55–57 (1991).
- Neilsen, J. B. *Adv. Inorg. Biochem.* **8**, 63–90 (1990).
- Trick, C. G., Andersen, R. J., Gillam, A. & Harrison, P. J. *Science* **219**, 306–308 (1983).
- Trick, C. G. *Curr. Microbiol.* **18**, 375–378 (1989).
- Andersen, R. J., Wolfe, M. S. & Faulkner, D. J. *Mar. Biol.* **27**, 281–285 (1974).
- Gauthier, M. J. & Flatau, G. N. *Can. J. Microbiol.* **22**, 1612–1619 (1976).
- Reid, R. T. & Butler, A. *Limnol. Oceanogr.* **36**, 1783–1792 (1991).
- Bax, A. & Summers, M. F. *J. Am. chem. Soc.* **108**, 2093–2094 (1986).
- Bronxli, P. & Gerig, J. T. *J. Am. chem. Soc.* **112**, 3719–3726 (1990).
- Teintze, M. & Leong, J. *Biochemistry* **20**, 6457–6462 (1981).
- Salituro, F. G., Agarwal, N., Hofman, T. & Rich, D. H. *J. med. Chem.* **30**, 286–295 (1987).
- Mayor, S. & Hosur, R. V. *Magn. Res. Chem.* **23**, 470–473 (1985).
- Chaberek, S. & Martell, A. E. *Organic Sequestering Agents* (Wiley, New York, 1959).
- Harris, D. C. *Quantitative Chemical Analysis* (Freeman, San Francisco, 1982).
- Loomis, L. D. & Raymond, K. N. *Inorg. Chem.* **30**, 906–911 (1991).
- Perrin, D. D. & Dempsey, B. *Buffers for pH and Metal Ion Control* (Chapman and Hall, London, 1974).
- Jalal, M. A. F. et al. *J. Am. chem. Soc.* **111**, 292–296 (1989).
- Takahashi, A. et al. *J. Antibiotics* **40**, 1671–1676 (1987).
- Marfey, P. *Carlsberg Res. Commun.* **49**, 591–596 (1984).
- Rich, D. H., Sun, E. T. O. & Ulm, E. J. *med. Chem.* **23**, 27–33 (1980).
- Schwarzenbach, G. & Schwarzenbach, K. *Helv. chim. acta* **46**, 1390–1400 (1963).
- Abdallah, M. A. et al. Abstracts XXVII Int. Conf. Coordination Chemistry, W20 (1989).
- Wong, G. B., Kappel, M. J., Raymond, K. N., Matzanke, B. & Winkelmann, G. *J. Am. chem. Soc.* **105**, 810–815 (1983).
- Anderegg, G., L'Epattier, F. & Schwarzenbach, G. *Helv. chim. acta* **46**, 1409–1422 (1963).
- Harris, W. R., Carrano, C. J. & Raymond, K. N. *J. Am. chem. Soc.* **101**, 2722–2727 (1979).
- Harris, W. R., Carrano, C. J. & Raymond, K. N. *J. Am. chem. Soc.* **101**, 2213–2214 (1979).
- Martell, A. E. & Smith, R. M. *Critical Stability Constants Vol. 1* (Plenum, New York, 1974).

ACKNOWLEDGEMENTS. We thank B. C. M. Potts for technical assistance and helpful discussions. A.B. is grateful for support from the NIH and the California Sea Grant. A.B. is an Alfred P. Sloan Research Fellow. R.T.R. is Sea Grant Trainee. D.J.F. is grateful for support from the California Sea Grant. D.H.L. acknowledges support from the NIH and NSF (towards purchase of the GN-500 spectrometer at Emory used in the heteronuclear experiments).

Cod spawning on a migration highway in the north-west Atlantic

George A. Rose

Science Branch, Department of Fisheries and Oceans, PO Box 5667, St John's, Newfoundland, Canada A1C 5X1

FIVE hundred years of fishing and fifty years of research have produced only vague accounts of the spawning and migrations of Atlantic cod (*Gadus morhua*) off Newfoundland in the north-west Atlantic¹. Here I report the discovery of 'spawning columns' and a 'highway' used by cod to traverse the northeastern Newfoundland Shelf during annual springtime feeding migrations. Sea research using echosounders^{2,3} showed that cod spawned in dense shoals (to one fish per m³) that featured midwater spawning columns comprised of pairs of fish. Immature joined mature post-spawning cod to migrate in large (scales of tens of kilometres and hundreds of millions of fish) size-structured aggregations led by larger 'scouts'. Cod traversed the cold waters of the shelf along a deep highway of warm oceanic water (2–2.5 °C). During migration, fish spacing appeared to maximize search volumes while maintaining visual contact. Aggregations fragmented when prey (capelin *Mallotus villosus*; shrimp, *Pandalus* spp.) were encountered.

Three shoreward cod migration highways across the north-

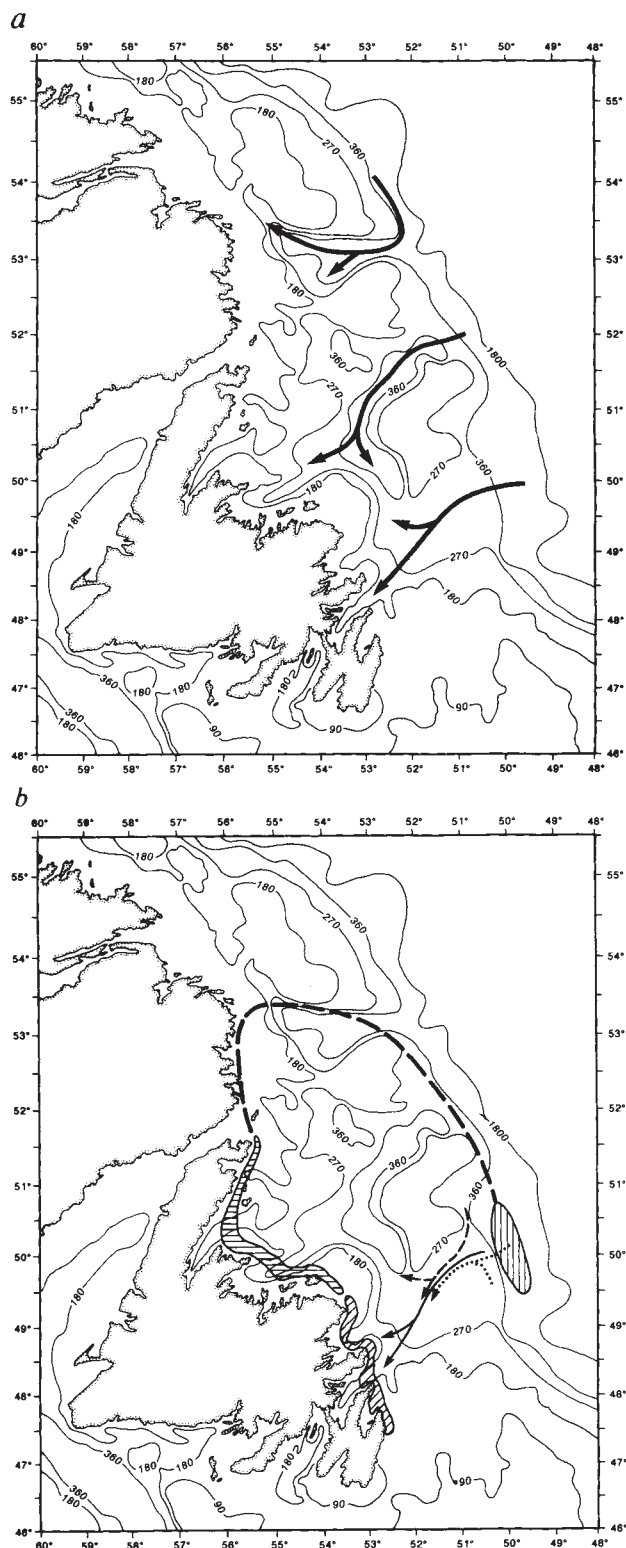


FIG. 1 a, The NE Newfoundland and Labrador shelves showing the cod migration routes predicted by simulations under an assumption that cod would follow warmer oceanic water masses from the offshore wintering to nearshore summer feeding areas. Isobaths in metres. b, Migration patterns of cod on the NE Newfoundland and Labrador shelves observed in June and July of 1990 (solid line), 1991 (broken line), and 1992 (dotted line) with echosounders. In each year fish were tagged along the route. Hatched areas represent nearshore areas of tag returns in July (diagonal hatching) and August (horizontal hatching) 1990 and 1991 and from offshore winter fisheries 1991 and 1992 (vertical hatching). Heavy broken line joining hatched areas outlines the boundary of the circular migration route indicated from these studies. Exact routes from inshore to offshore is not known.

eastern Newfoundland Shelf were predicted under an assumption that routes would follow trenches holding warm oceanic waters (2–3 °C) undercutting colder shelf waters (<0 °C) (Fig. 1a). Cod travelled the southerly highway in the springs of 1990, 1991 and 1992 (Fig. 1b). No shoreward migration was observed elsewhere. Returns of tagged cod indicated cod struck the coast at the shoreward end of the highway then dispersed northwards during the summer. Tag returns also suggested cod completed an autumn return migration⁴ to the wintering areas at the offshore margins of the highway.

Cod migrated shorewards in large aggregations of mature fish accompanied by juvenile four- and five-year-olds (>40 cm long). The largest aggregation was observed in the highway in 1990 (Fig. 2). When discovered, it occupied an area of $16.4 \times 10^8 \text{ m}^2$ and a volume of $4.1 \times 10^{10} \text{ m}^3$ where sea temperatures were 2–2.5 °C (Fig. 3). Acoustic survey densities² ranged up to 0.8 fish per m^3 (mean, 0.014; s.e., 0.001; $n=3,174$; mean target strength of –33.0 decibels per fish), indicating a total abundance of 5.7×10^8 fish (>80% of stock estimate⁵). Smaller aggregations were observed in 1991 and 1992. Cod traversed the outer shelf along the 2 °C isotherm until food was encountered. Then migration routes were modified as cod pursued prey (Fig. 3). In each year aggregations remained intact until abundant food (capelin and shrimp) was encountered.

The formation of a migration aggregation was observed in 1992. Its kernel was a spawning shoal initially observed towards the seaward margins of the highway. This shoal remained relatively immobile (daily displacements of 1–2 km) during ten days of observation. It was shaped like an inverted saucer, with radius 1.2 km and centre height 40–45 m (estimated volume, $1.1 \times 10^8 \text{ m}^3$) (Fig. 4a). This aggregation illustrates properties typical of spawning aggregations observed in this study: uniformly high densities (to >1 fish per m^3), very discrete edges, and daytime spawning columns consisting of multiple pairs of fish above the main group. When discovered, half of the mature females were about to spawn, and within 3–4 weeks >95% had spawned. As fish in the spawning school dispersed somewhat (Fig. 4b), smaller and immature cod migrating in from deeper waters joined them to form the migration aggregation (Fig. 4c). As migration started, aggregation dimensions expanded in three dimensions (to 20 km by 50–150 m) as fish moved forwards and

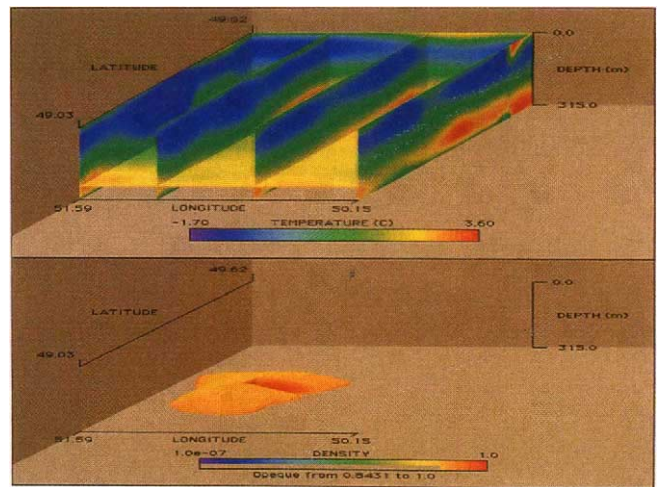


FIG. 2 Thermal structure across the migration highway in 1990 and the core of the cod aggregation while at rest. Temperatures range from –1.7 °C (dark blue) in upper Arctic waters to 3.6 °C (bright red) in deeper Atlantic Ocean waters of the migration highway. Virtually all cod located where temperatures were 2.0–2.7 °C. Images created by linear interpolations of acoustic fish density (38 kHz, 100 m horizontal by 10 m vertical resolution) and temperature from 7 transects across area within 24 h (21 bathythermograph profiles recorded). Fish aggregation view cut out to show internal density structure. Scale of fish density logarithmic (actual arithmetic range of densities shown is 0.08 to 0.8 fish per m^3).

upwards (Fig. 4d). Acoustic calculations indicated that on average, cod occupied 66–83 m^3 of sea water during migration (4.0–4.5 m, or 8–10 body lengths from their nearest neighbour). In contrast, mature cod in the spawning school occupied on average only 1 m^3 of seawater (one body-length spacing).

Mean migration aggregation densities did not differ between years (0.014, 0.012 and 0.15 fish per m^3 in 1990, 1991 and 1992, respectively). At 350 m in these waters in daytime, light intensi-

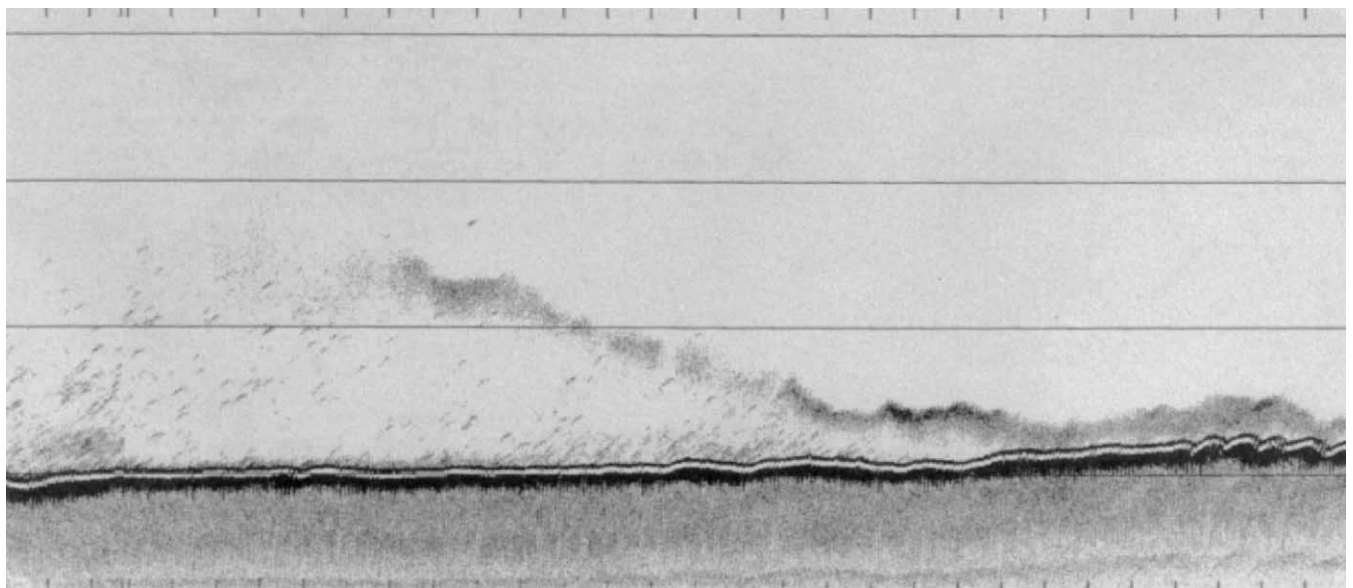


FIG. 3 Echogram of cod scouts advancing from right to left and encountering capelin 23 June 1991 at 03:30 h along the migration highway. Capelin are a dark cloud rising off bottom; cod are single traces rising

beneath the capelin. Total water depth, 350 m; horizontal grids, 50 m (only bottom waters shown); total horizontal view, 4.0 km.

ties are of the order of 10^{-2} to $10^{-3} \mu\text{W m}^{-2}$ (refs 6–8). In this environment cod may discriminate light intensity increments of 10% (Weber fractions)^{8,9}; cod probably show a low degree of visual contrast (0.2)¹⁰, hence one cod might detect another at distances of 7 m. These calculations suggest cod space themselves during migration at intervals near but not beyond visibility lim-

its. Such spacing would maximize the area swept for prey while maintaining visual contact (Fig. 3).

Migrating cod aggregations were structured by fish size. In 1990 and 1991 migrations appeared to be led by scouts (Fig. 2). Acoustic and trawl data suggested that scouts were larger fish (lengths over 50 cm). In 1992, systematic sampling of the moving

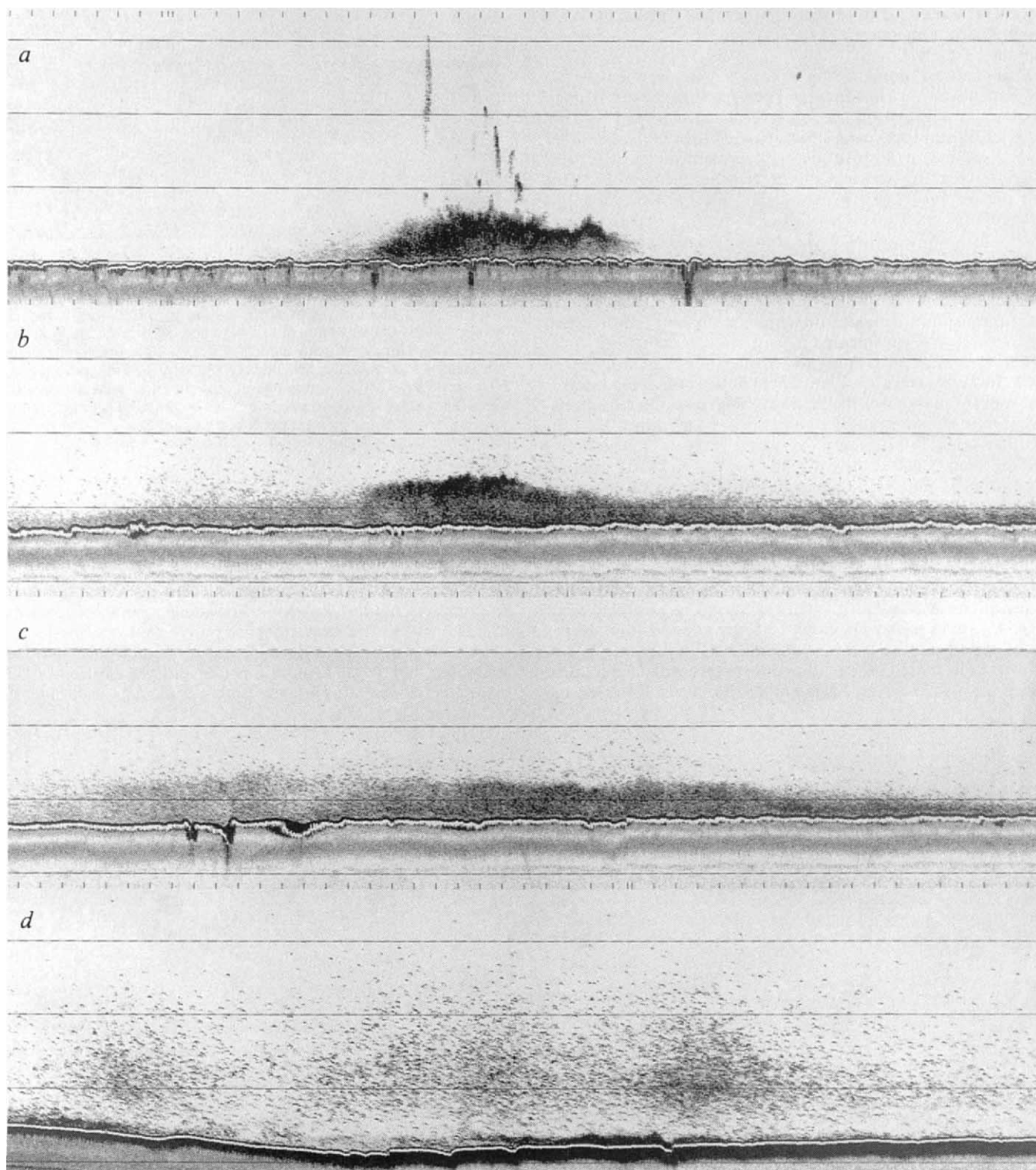


FIG. 4 Echograms of cod in the migration highway in the spring of 1992. All panels share common features: total water depth, 350–375 m; horizontal grids, 50 m (only bottom waters shown); total horizontal view, 7.5 km; dense masses of fish appear as black, with lower densities greyer; single fish are smaller traces. *a*, The spawning school that formed the kernel of the migration aggregation and its characteristic

columns (maximum internal school densities ~ 1 fish per m^3), on 11 June at 13:00 h. *b*, The first evidence of some dispersal of the spawning school, on 14 June at 03:00 h. *c*, Onset of pre-migratory behaviour and joining up with smaller fish, on 15 June at 22:00 h. *d*, The migration formation, with single fish well spaced and off the bottom, on 9 July at 22:00 h.

aggregation by a trawler directed from an acoustic vessel confirmed that scouts were the largest fish (median, 50 cm for scouts, 46 cm for middle; 44 cm for rearguard). This result occurred despite a low range of sizes in 1992 (interquartile ranges 45–57 cm in 1990, 41–53 cm in 1991 and 42–48 cm in 1992). Rearguard fish apparently followed the scouts. On several occasions scouts veered sharply off-course in pursuit of prey (Fig. 3). In each case the whole aggregation veered: it is unlikely that the rearguard could have seen the prey initially and were probably following the fish ahead.

Several questions emerge from these observations. How do cod achieve a high degree of spatial precision in their migrations? What signposts do they use? Do cod pilot¹¹ relative to persistent hydrographic and bathymetric features (the highway was predicted on the basis of such a thermal assumption). Indeed, the constancy of the temperatures chosen by cod (when they could go elsewhere) suggests that routes develop and persist, at least in part, because of suitable and stable temperatures. Still, the mechanisms by which cod migrate await further inquiry. The lack of use of the northerly highways is attributed to the southerly cod distributions of recent years¹².

What is the function of the spawning columns? In laboratory experiments, cod spawn in pairs in midwater after more demersal courtship¹³. Acoustic records suggest the columns may comprise pairs of spawning fish, having paired near bottom then swum up to spawn.

Why do migrating cod form large aggregations^{4,14}? Could the answer relate to feeding and learning by juveniles? These fish migrate long distances in search of food. Perhaps younger follow older cod to take advantage of feeding opportunities, and by doing so learn migration routes^{15–17}? Juvenile Arcto-Norwegian cod have been reported to make a dummy run accompanying adults to the spawning grounds^{18,19}. If routes are partially learned then loss of older fish might lead to changed routes. Interestingly, a major reduction of older fish in 1991 and 1992 paralleled lower rates of passage shoreward than in 1990 (but equivalent daily displacements). Perhaps older fish sustain migratory routes and behaviours. The migration patterns of Norwegian spring spawning herring (*Clupea harengus*) changed after adult stocks were depleted in the late 1960s (ref. 20). The original patterns have yet to be re-established. For Newfoundland cod, migration patterns may also be expected to change, possibly irrevocably, in response to population declines in the early 1990s. □

Received 8 July; accepted 8 October 1993.

- Harris, L. *Independent Review of the Northern Cod Stock* (Comm. Dir. Dept Fish. Oceans, Ottawa, 1990).
- Sund, O. *Nature* **135**, 953 (1935).
- MacLennan, D. N. & Simmonds, E. J. *Fisheries Acoustics* (Chapman & Hall, London, 1992).
- Cushing, D. H. & Harden-Jones, F. R. *Nature* **218**, 918–920 (1968).
- Baird, J. W., Bishop, C. A., Brodie, W. B. & Murphy, E. F. *CAFSAC Res. Doc.* 92/75 (1992).
- Blaxter, J. H. S. in *Marine Ecology* (ed. Kinne, O.) 213–320 (Wiley, New York, 1970).
- Clarke, G. L. & Denton, E. J. in *The Sea* (ed. Hill, M. N.) 456–468 (Wiley, New York, 1962).
- Clarke, G. L. in *Proceedings of the International Symposium on Biological Sound Scattering in the Ocean*, 1970 (ed. Farquhar, G. B.) 41–50 (Dept of the Navy, Washington DC, 1971).
- Blaxter, J. H. S. in *Fish Behavior and its Use in the Capture and Culture of Fishes* (eds Bardach, J. E., Magnuson, J. J., May, R. C. & Reinhart, J. M.) 33–55 (ICLARM Conf. Proc. 5) (Manila, Philippines, 1980).
- Duntley, S. Q. in *The Sea* (ed. Hill, M. N.) 452–455 (Wiley, New York, 1962).
- Griffin, D. R. in *Recent Studies in Avian Biology* (ed. Wolfson, A.) 154–197 (Univ. Illinois Press, Urbana, 1955).
- de Young, B. & Rose, G. A. *Can. J. Fish. Aquat. Sci.* (in the press).
- Brawn, V. M. *Behaviour* **XVIII**, 177–198 (1961).
- Pitcher, T. J. in *The Behaviour of Teleost Fishes* (ed. Pitcher, T. J.) 294–337 (Johns Hopkins, Baltimore, 1986).
- Kieffer, J. D. & Colgan, P. W. *Rev. Fish Biol. Fisheries* **2**, 125–143 (1992).
- Dodson, J. J. *Env. Biol. Fishes* **23**, 161–182 (1988).
- Trout, G. C. *Fish. Invest. London II*, XXI (6) 51 (1957).
- Woodhead, A. D. *J. mar. Bio. Ass. U.K.* **38**, 417–422 (1959).
- Harden-Jones, F. R. *Fish Migration* (Arnold, London, 1968).
- Dragesund, O., Hamre, J. & Ulltang, O. *Rapp P.-V. Reun. Cons. Int. Explor. Mer.* **177**, 43–71 (1980).

ACKNOWLEDGEMENTS. I thank D. Porter, M. Bloom and Hydroacoustic and Groundfish staff at the Northwest Atlantic Fisheries Centre for assistance, and G. Mertz and J. Morgan for comments on the manuscript. Funding was provided by the Ocean Production Enhancement Network (Government of Canada), and the Department of Fisheries and Oceans Northern Cod Science Program.

A *C. elegans* mutant that lives twice as long as wild type

Cynthia Kenyon, Jean Chang, Erln Gensch, Adam Rudner & Ramon Tabtiang

Department of Biochemistry and Biophysics, University of California at San Francisco, San Francisco, California 94143-0554, USA

WE have found that mutations in the gene *daf-2* can cause fertile, active, adult *Caenorhabditis elegans* hermaphrodites to live more than twice as long as wild type. This lifespan extension, the largest yet reported in any organism¹, requires the activity of a second gene, *daf-16*. Both genes also regulate formation of the dauer larva, a developmentally arrested larval form that is induced by crowding and starvation and is very long-lived^{2–4}. Our findings raise the possibility that the longevity of the dauer is not simply a consequence of its arrested growth, but instead results from a regulated lifespan extension mechanism that can be uncoupled from other aspects of dauer formation. *daf-2* and *daf-16* provide entry points into understanding how lifespan can be extended.

Temperature-sensitive mutations in several genes lead to formation of long-lived dauer larvae in the absence of inducing signals. To learn whether these mutations might also increase lifespans of non-daughters, we examined the lifespans of dauer-constitutive *daf-2*, *daf-11*, *daf-7* and *daf-14* mutants (refs 4, 5 and E. Malone and J. Thomas, personal communication). The decision to enter the dauer stage, an alternative third larval stage (L3) must be made during the first larval stage (L1). Therefore, we shifted the mutant larvae to the non-permissive temperature at later stages, allowed them to grow to adulthood, and measured their lifespans⁶. Because many temperature-sensitive mutations lead to reduced gene activity even at the permissive temperature, we also measured the lifespans of these mutants at lower temperatures. The lifespans of *daf-11*, *daf-7* and *daf-14* adults were not noticeably different from wild type (Fig. 1 legend). In contrast, each of three *daf-2* mutations, *e1370*, *sa189* and *sa193* greatly extended lifespans of adults when shifted to the non-permissive temperature during L4 (Fig. 1a, b legend). *daf-2(e1370)* and *daf-2(sa193)* were also examined at the permissive temperature; the lifespans of both were extended (Fig. 1c, and data not shown).

The magnitude of the lifespan extension was striking. For example, when shifted from 15 °C to 20 °C at the L4 stage, the mean lifespan of wild type was 18 days, whereas that of *daf-2(sa189)* was 42 days (Fig. 1a), 2.3-fold greater than wild type. The difference between ageing rates of *daf-2* and wild-type animals was also readily apparent. When all the wild-type animals had died or become immobile, 90% of the *daf-2(e1370)* animals still moved actively on their plates (Fig. 2c). The long-lived *daf-2* adults resembled wild-type adults much more closely than the dauer. The dauer is developmentally arrested as a small, thin, and sexually immature larval form⁴. In contrast, the *daf-2* non-dauer animals became full-size adults, and appeared to feed and defaecate normally (Fig. 2, and data not shown). Furthermore, at 20 °C, where its lifespan was twice that of wild type, *daf-2(e1370)* animals had nearly normal brood sizes (212 ± 36 for *daf-2(e1370)* ($n=29$); as compared to 278 ± 35 for wild-type animals grown in parallel ($n=19$)).

Because one theory of ageing postulates that reproduction decreases lifespan¹, we wondered whether the slight decrease in the brood size of *daf-2* animals could possibly be responsible for their longevity. *spe-26* hermaphrodites, which are defective in spermatogenesis, live longer than the wild type⁷. This is consistent with the model that sperm production decreases the lifespans of *C. elegans* hermaphrodites¹. To investigate directly whether reproduction shortens lifespan, we ablated the precursors of the germ cells and somatic gonad (the cells Z1, Z2, Z3 and Z4) in wild-type hermaphrodites using a laser microbeam^{8,9}. The

lifespans of these animals were normal (Fig. 3a), as were the lifespans of *fem-3(e1996)* mutants¹⁰, which do not produce sperm or self-progeny (Fig. 3b). We conclude that production of germ cells and progeny in itself does not shorten the lifespan of the *C. elegans* hermaphrodite. As expected, *daf-2(e1370)* animals still exhibited the longevity phenotype after ablation of the gonad and germ cells (data not shown). It seems likely that factors other than decreased germ cell or progeny production are responsible for the longevity of *C. elegans* lifespan mutants, including *spe-26*.

Why did dauer-constitutive mutations in the genes *daf-7*, *daf-11* and *daf-14* not affect lifespan? These genes function upstream of *daf-2* in regulating dauer formation⁵. They appear to act within partially redundant parallel pathways to keep *daf-2* activity levels high in the absence of dauer-inducing signals¹¹. *daf-11* lies in one pathway, and *daf-7* and *daf-14* lie in the other pathway. Our data imply that mutations in these genes do not reduce the activity of *daf-2* in such a way that the lifespans of adults can be lengthened. It may be that because these gene functions are partially redundant, none of these mutations alone has a strong enough effect on *daf-2* levels to affect lifespan. Another possibility is that *daf-11*, *7* and *14* are capable of regulating *daf-*

2 levels only during L1, when dauer formation can be initiated. If so, shifting these mutants to the non-permissive temperature at later times would have no effect on *daf-2* levels.

How do *daf-2* mutations extend lifespan? The gene *daf-16* functions downstream of *daf-2* to promote dauer formation (refs 5, 12 and D. L. Riddle, personal communication). We found that the *daf-16* mutation *m26* completely suppressed the lifespan extension of *daf-2(e1370)* adults (Fig. 4). This suggests that *daf-2* mutations extend lifespan by allowing inappropriate activation of *daf-16* function. But how *daf-16* activity extends lifespan is unclear. The simplest model is that *daf-16* activity extends the lifespans of *daf-2* adults by triggering expression of a regulated lifespan extension mechanism that is normally coupled to dauer formation. This would imply that the long lifespan of the dauer is not simply a consequence of the fact that it is developmentally arrested, as has been widely assumed^{4,13}. Alternatively, the mechanism by which *daf-16* extends the lifespans of *daf-2* adults could be different from the mechanism that allows dauers to live longer.

Could a similar type of lifespan extension operate in other organisms? Lifespans of mammals, and, to a limited extent, *C. elegans*, can be increased by food limitation¹⁴. It is possible that

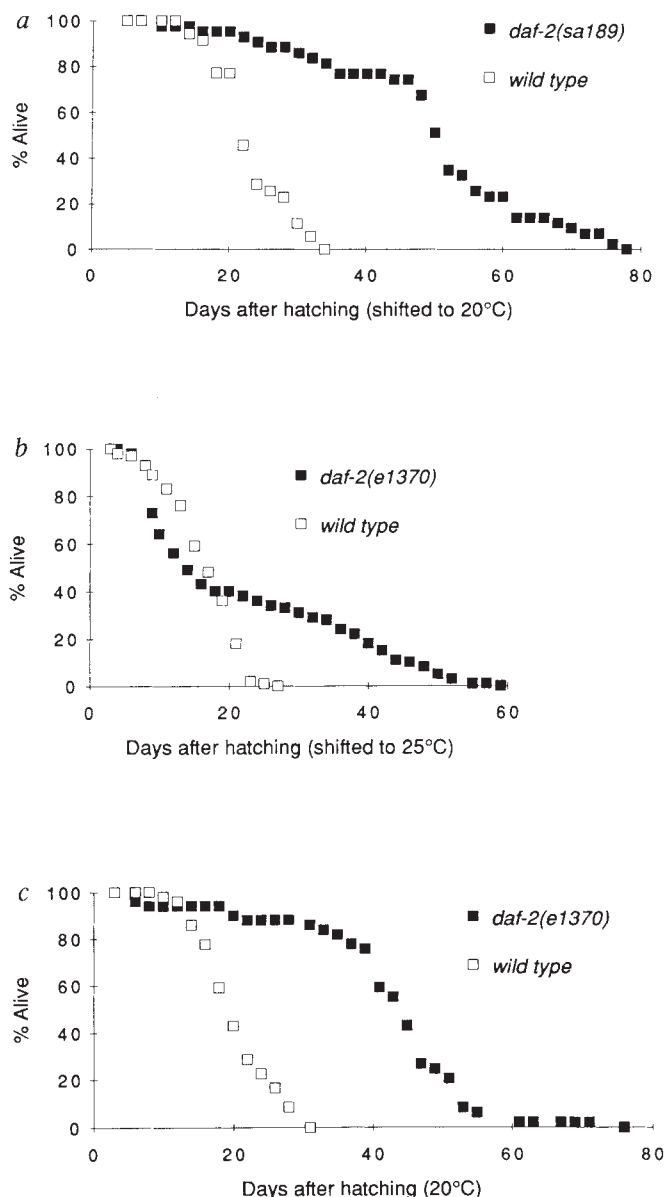


FIG. 1 *daf-2* adults have extended lifespans. a, *daf-2(sa189)* and wild type (N2) shifted from 15 °C to 20 °C, the non-permissive temperature, at the L4 stage. Mean lifespans were 42 days for *sa189* ($n=43$) and 18 days for N2 ($n=35$). Comparisons of these means with each other using the log-rank test¹⁷ yielded a χ^2 value of 62.9 (P , the probability of these lifespan differences occurring by chance, $\ll 0.00001$). b, Lifespans of *daf-2(e1370)* and N2 (wild-type) animals shifted from 20 °C to 25 °C, the non-permissive temperature, as L4 larvae. Mean lifespans were 22.5 days for *e1370* ($n=120$) and 17 days for N2 ($n=119$). $\chi^2=10.0$ ($P<0.0016$). This is one of two experiments; both gave similar results. When shifted to 25 °C, about half the *daf-2(e1370)* animals died as young adults containing newly hatched progeny. c, Lifespan extension of *daf-2(e1370)* cultured continuously at 20 °C. Mean lifespans were 42 days for *daf-2(e1370)* ($n=42$) and 20 days for N2 ($n=49$); $\chi^2=67.1$ ($P<0.00001$). At this temperature, *daf-2(e1370)* mutants did not die prematurely as 'bags of worms', which probably contributed to the greater mean lifespan of the population. The third allele, *daf-2(sa193)*, also exhibited an extended lifespan when shifted from 20 °C to 25 °C (non-permissive temperature). Mean lifespan for *sa193* was 25 days ($n=21$) and for N2 controls, 15 days ($n=44$). The efficiency of dauer formation in this strain was lower than *daf-2(sa189)* and *daf-2(e1370)* (not shown), as was its maximum lifespan (39 days). When cultured continuously at 20 °C, the mean lifespan of this strain was also extended slightly (not shown). *sa189* and *sa193* were isolated recently; the N2 control used in those lifespan experiments was the direct parent of these mutants. *sa189* and *sa193* were isolated recently by E. A. Malone and J. Thomas (personal communication). The lifespans of *daf-7(e1372)*, and *daf-14(m77)* were examined when grown continuously at 15 °C, or when shifted from 15 to 20 °C and from 15 to 25 °C as L4 animals. The lifespan of *daf-11(m84)* was determined after shifting L4 animals from 15 to 25 °C during L4. Under some conditions, a fraction of the mutants died prematurely as bags of worms¹¹. The lifespans of the remaining animals were not significantly different from wild type (data not shown). In all experiments, five L4 animals were placed on NGM plates seeded with *Escherichia coli* OP50 (ref. 18). Animals were transferred to fresh seeded plates every two days when producing progeny, and about once a week thereafter. Animals were scored as dead when they no longer moved or twitched when prodded or pushed several times. All experimental strains were grown in parallel with wild type.

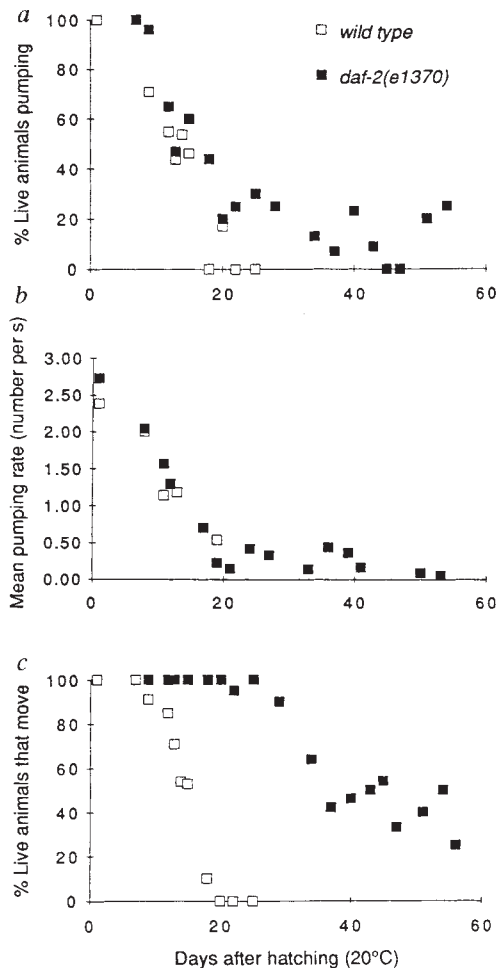
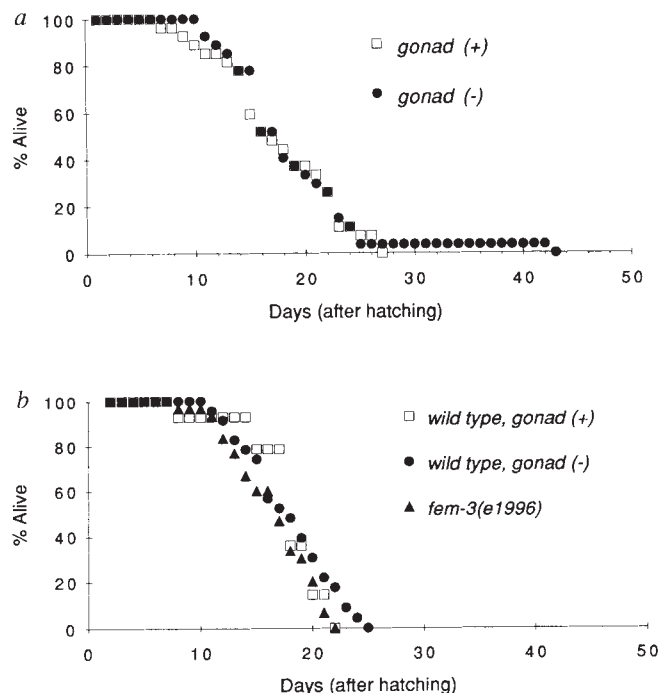


FIG. 2 Behaviours of *daf-2(e1370)* and wild-type adults. Animals were cultured individually on separate plates at 20 °C, and their feeding behaviour and movement were observed ($n=19$ for N2 and $n=20$ for *daf-2(e1370)*). *C. elegans* ingests food by periodic contraction of its pharynx, called 'pumping'¹⁹. To determine the per cent of live animals pumping (a), each animal was examined twice within a 1 h period for at least 30 s each time. Average pumping rate (b) was determined by measuring the pumping rates of at least five randomly selected animals that were pumping (dead and non-pumping animals were not included). As both wildtype and *daf-2* mutants aged, their pumping became sporadic. Often the animals lay still when first observed, but when stimulated, began to move across the plates, and often to pump. Individuals not pumping on one day would sometimes be pumping a few days later. Thus it seems likely that whereas young adults pump continuously, older animals (including mutant and wild type) pump intermittently until close to death. It is possible that the decrease in pumping rate, which occurs with similar kinetics in wild-type and *daf-2* animals, is not due to senescence, but instead reflects a decreased requirement for food when production of progeny ceases. To assay movement (c), we investigated what fraction of animals were moving actively on their plates, or began to move across the plates when touched with an eyelash²⁰. Animals that moved their heads or bodies when stimulated but stayed in place were scored as negative. As *C. elegans* senesces^{2,3,21}, it becomes lethargic and eventually moves only its nose when touched. The inability to move appears to be a good indicator of impending death for both wild type and *daf-2*. (Compare the survival curves of these same animals, shown in Fig. 4, with the per cent capable of moving.) *daf-2* mutants did appear different from wild-type adults in at least two respects. First, they grew to adulthood more slowly than wild type, taking 3 days rather than 2 at 20 °C. In addition, like the dauer, their intestines appeared dark when viewed with a dissecting microscope. Neither of these phenotypes is sufficient to account for the extended lifespans of *daf-2* animals, as *daf-7* and *daf-14* mutants exhibit these phenotypes¹¹ but did not have extended lifespans.

FIG. 3 The lifespan of the *C. elegans* hermaphrodite is not affected by the production of germ cells or progeny. a, Lifespans of hermaphrodites lacking the reproductive system. The germ-cell precursors, Z2 and Z3, as well as the precursors of the somatic gonad, Z1 and Z4, were ablated in 27 newly hatched N2 hermaphrodites using standard procedures^{8,9,22}. Mean lifespans were 18 days (s.d. = 6) for ablated animals ($n=27$) and 17 days (s.d. = 5) for N2 controls ($n=27$). These survival curves were not significantly different from one another ($\chi^2=0.015$; $0.89 < P < 0.92$). Control animals were anaesthetized with 10 mM sodium azide and placed on agar pads in parallel with experimental animals. To ensure that the ablation procedure was satisfactory, a sample of 5 operated animals was examined using Nomarski optics several hours after the microsurgery, and found to lack Z1–Z4. In addition, all the experimental animals were sterile and appeared to lack the gonad when viewed with a dissecting microscope. The act of ablation itself was unlikely to affect lifespan, because *daf-2* mutants still expressed the full longevity phenotype following ablation (data not shown). b, Lifespans of *fem-3(e1996)* mutants, which develop as females¹⁰. Germ-line precursors that normally differentiate into sperm differentiate into oocytes instead. In parallel with these animals, we examined a second set of animals in which Z1–Z4 had been ablated. This experiment demonstrates the reproducibility of the gonad ablation experiment. Mean lifespans were 16 days for *fem-3(e1996)* ($n=30$); 17 days for gonad-ablated animals ($n=23$); and 17 days for N2 controls ($n=14$). These survival curves were not significantly different from one another: $\chi^2=0.01$ and $0.89 < P < 0.93$ for gonad (+) versus (–), and $\chi^2=0.94$ and $0.32 < P < 0.34$ for gonad (+) versus *fem-3*.



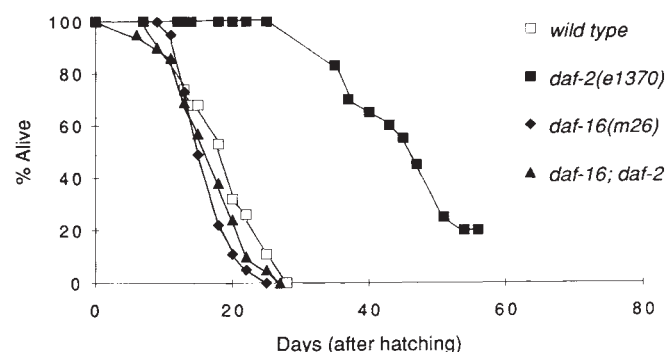


FIG. 4 The longevity of *daf-2* mutants is blocked by *daf-16(m26)*, a dauer-defective mutation in a gene that acts downstream of *daf-2* in dauer formation. Lifespans were determined in parallel for N2, *daf-2(e1370)* and *daf-16(m26)* single mutants as well as *daf-16(m26); daf-2(e1370)* double mutants. Mean lifespans were 17 days for m26 ($n=37$); 17 days for m26; e1370 ($n=42$); and 19 days for N2 ($n=19$). For *daf-2(e1370)*, $n=20$. The lifespan differences between N2, *daf-16* and *daf-16; daf-2* were not significantly different from one another (for example, for N2 versus *daf-16*, $\chi^2=1.68$; $0.19 < P < 0.21$), whereas differences between these three strains and *daf-2* were statistically significant (for example, for *daf-2* versus *daf-16; daf-2*, $\chi^2=50.6$; $P < 0.00001$).

daf-2 mutations elicit an internal signal also generated by food limitation, which can extend lifespan without substantially affecting fertility or general activity. It would be interesting to learn whether lifespan extension caused by food limitation in *C. elegans* requires *daf-16* activity. Strains of *Drosophila* with mean lifespans roughly 30% longer than normal have been obtained by selective breeding^{1,15}. Like *daf-2* mutants, these flies appear robust and healthy. Intriguingly, they are also resistant to starvation and desiccation¹⁶, as is the *C. elegans* dauer. Perhaps a common underlying mechanism operates to extend the lifespans of these different types of organisms. If so, then identification of *C. elegans* genes that act downstream of *daf-16* could lead to a general understanding of how lifespan can be extended. □

Received 13 May; accepted 15 September 1993.

1. Finch, C. E. *Longevity, Senescence and the Genome* (Chicago Press, Chicago, 1990).
2. Zuckerman, B. M. *Nematodes as Biological Models* (Academic, New York, 1980).
3. Johnson, T. E. & Wood, W. B. *Proc. natn. Acad. Sci. U.S.A.* **79**, 6603–6607 (1982).
4. Riddle, D. L. *The Nematode Caenorhabditis elegans* 393–414 (Cold Spring Harbor Laboratory Press, New York, 1988).
5. Vowles, J. J. & Thomas, J. H. *Genetics* **130**, 105–123 (1992).
6. Swanson, M. & Riddle, D. *Devi Biol.* **84**, 27–40 (1981).
7. Van Voorhies, W. A. *Nature* **360**, 456–458 (1992).
8. Sulston, J. & White, J. *Devi Biol.* **78**, 577–597 (1980).
9. Kimble, J. *Devi Biol.* **87**, 286–300 (1981).
10. Hodgkin, J. *Genetics* **114**, 15–52 (1986).
11. Thomas, J. H., Birnby, D. A. & Vowles, J. J. *Genetics* **134**, 1105–1117 (1993).
12. Riddle, D. L., Swanson, M. M. & Albert, P. S. *Nature* **290**, 668–671 (1981).
13. Klass, M. R. & Hirsch, D. *Nature* **280**, 523–525 (1976).
14. Klass, M. R. *Mechanisms Ageing Dev.* **6**, 413–429 (1977).
15. Rose, M. R. *Can. J. Zool.* **62**, 1661–1667 (1984).
16. Service, P. M., Hutchinson, E. W., Mackinley, M. D. & Rose, M. R. *Physiol. Zool.* **58**, 380–389 (1985).
17. Woolson, R. *Statistical Methods for the Analysis of Biomedical Data* ch. 14 (Wiley, New York, 1987).
18. Sulston, J. *The Nematode Caenorhabditis elegans* 587–606 (Cold Spring Harbor Laboratory Press, New York, 1988).
19. White, J. *The Nematode Caenorhabditis elegans* 81–122 (Cold Spring Harbor Laboratory Press, New York, 1988).
20. Sulston, J., Dew, M. & Brenner, S. *J. comp. Neurol.* **163**, 215–226 (1975).
21. Johnson, T. E. *Proc. natn. Acad. Sci. U.S.A.* **84**, 3777–3781 (1987).
22. Waring, D. A. & Kenyon, C. *Cell* **60**, 123–131 (1990).

ACKNOWLEDGEMENTS. We thank J. Thomas, D. Riddle, S. Barondes and members of the Kenyon laboratory for helpful discussions and for improving the manuscript. We especially thank J. Thomas for sending us unpublished alleles of *daf-2*, *sa193* and *sa189*, as well as their direct parent N2 strain, D. Riddle for sending *daf-16(m26)* and *daf-16(m26); daf-2(e1370)* and C. Bargmann for giving us *daf-2(e1370)*. This work was supported by a Packard Foundation grant to C.K.; A.R. and R.T. were supported by Howard Hughes Medical Institute graduate student fellowships, and J.C. and E.G. were supported by NIH graduate student training grants.

Modulation of the cell cycle contributes to the parcellation of the primate visual cortex

Colette Dehay, Pascale Giroud, Michel Berland*, Iain Smart† & Henry Kennedy‡

INSERM U371, Cerveau et Vision, 18 Avenue Doyen Lépine, 69500 Bron, France

* Faculté de Médecine Lyon-Nord, Hôpital Claude Bernard, Service de Gynécologie et Obstétrique, Oullins 69600, France

† Department of Anatomy and Physiology, University of Dundee, Dundee DD1 4HN, UK

AN as-yet unresolved issue in developmental neurobiology is whether the discrete areas that form the mammalian cortex emerge from a uniform cortical plate or whether they are already specified in the germinal zone^{1,2}. A feature of the primate striate cortex is that the number of neurons per unit area is twice that of anywhere else in the cerebral cortex³. Here we take advantage of this unique structural feature to investigate whether the extra striate cortical cells are due to increased neuron production during neurogenesis. We labelled precursors undergoing terminal cell division with ³H-thymidine and allowed them to migrate to the cortical plate. Cell counts revealed that their rate of production in the germinal zone of striate cortex is higher than in that giving rise to extrastriate cortex. Also, we used ³H-thymidine pulse injections to investigate cell cycle dynamics and found that this phase of increased production of striate cortical cells is associated with changes in the parameters of the cell cycle. These results show that cortical area identity is at least partially determined at the level of the ventricular zone.

A total of eight ³H-thymidine pulse injections have been carried out on fetal monkeys. Fetuses received an intraperitoneal injection of ³H-thymidine at known gestational ages and were returned to the uterus for a survival period before being processed for autoradiographic histological observation. The position and density of labelled cells were determined in either the cortical plate or the underlying germinal zone (Fig. 1a). Using frequent sampling, we were able to employ a statistical test (ANOVA) to show that differences between striate and extrastriate cortex were not a result of spurious variations in labelling density in each area; we found significant differences in labelling for all ages except in the E64 case (see Table 1). Experimental cases fell into two groups (I and II in Table 1) according to the length of the survival period.

In the first group (5 cases), a long survival period of 14–84 days allowed neurons to migrate out to the cortical plate as well as a variable degree of cortical differentiation. In this group the numbers of labelled cells per unit area of cortical plate were estimated in striate and extrastriate cortex. In all cases, regardless of the age of the fetus at the moment of injection, striate cortex was found to have significantly more labelled cells per unit area than adjacent extrastriate cortex. Quantitative analysis of the results showed that there were developmental changes in the percentage increase of labelled cells in striate cortex compared with extrastriate cortex (Table 1).

Differences in labelling of striate and extrastriate cortex are illustrated in Fig. 1b, which shows cortical labelling in a neonate that had received a ³H-thymidine pulse injection on embryonic day 81 (E81). Different intensities of labelling broadly correspond to successive generations of cells, because there is roughly a halving of labelling intensity at each mitosis⁴. The radial location of successive generations of cells can be determined from the positions of cells with maximum, 50% of maximum and 25% of maximum labelling. Besides the increased rate of cell production in striate cortex (as indicated by the higher densities of labelled neurons in this area), Fig. 1b illustrates two common

‡ To whom correspondence should be addressed.

features seen at all ages. First, the fact that heavily labelled cells tend to be at deeper locations than more lightly labelled cells confirms the inside-first, outside-last sequence of cortical neuron production⁵. Second, labelled cells in extrastriate cortex are slightly nearer the surface in the supragranular layers than those in striate cortex, confirming that cell production in striate cortex lags behind that in extrastriate cortex⁶.

The ages for ³H-thymidine injection were chosen to span the period of neuron production of infra- and supragranular layers⁷. The developmental age of the fetus at the time of injection strongly influenced the increase in labelling of striate cortex (Fig. 2). The youngest fetus at the time of injection was E66; this showed an increase in labelling of striate cortex of 7%. Increases were larger in older fetuses, reaching a maximum of 150% in the fetus injected on day E78. At later stages of cortical neuron production, the percentage increase in labelled striate neurons fell and the two fetuses injected on E81 gave values of 70 and 100%. These results show that increased rates of striate cortex neuron production are much lower at early stages of neurogenesis when infragranular layers are being generated. Increased striate cortex neuron production peaks at about E78 when superficial layers are being generated⁷.

The second short-survival group (3 cases; Table 1) enabled us to examine ³H-thymidine labelling in the germinal zone lining the ventricles and to determine whether the higher cell production in striate cortex compared to the same area of adjacent cortex was due to changes in cell cycle parameters. The percentage of tritiated cells relative to the total population of germinal cells is referred to as the labelling index and reflects the proportion of precursor cells in S phase (the phase of DNA synthesis

TABLE 1 Experimental cases and statistical analysis of labelling differences of striate and extrastriate cortex

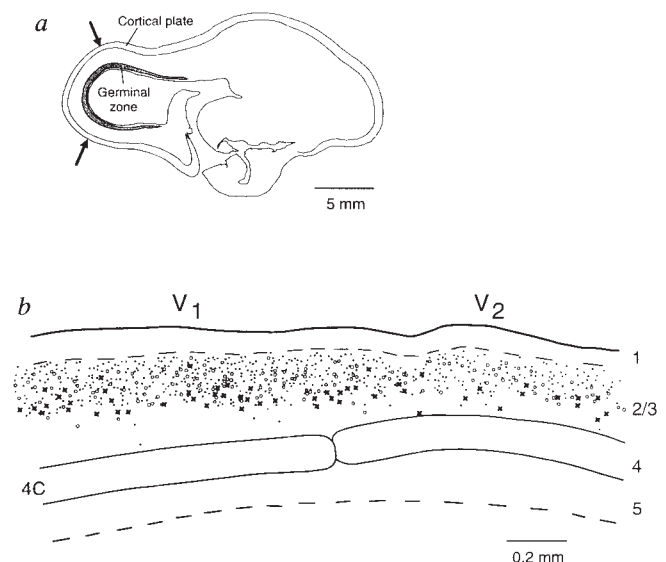
Group I		Differences of labelling density in cortex	
Age at ³ H-thymidine injection	Age at perfusion		
E66	E121	$F=8.28$	$P=0.0048$
E72	E94	$F=26.75$	$n=4,418$ $P=0.0001$
E78	E92	$F=40.39$	$n=2,458$ $P=0.0001$
E81*	E110	$F=105.82$	$n=12,919$ $P=0.0001$
E81	E165	Unpaired Student <i>t</i> -test	$n=2,505$ $P=0.0001$ $n=5,448$
Group II		Differences of labelling index in the germinal zone	
Age at ³ H-thymidine injection and perfusion	Labelling index (%)		
E64	A17, 20.15 A18, 20.38	$F=1.21$	$P=0.3029$ $n=13,223$
E78	A17, 29.65 A18, 19.88	$F=11.38$	$P=0.0097$ $n=19,006$
E94	A17, 6.97 A18, 5.41	$F=14.76$	$P=0.0049$ $n=14,220$

n, Total number of counted neurons; E represents the embryonic indicated by the number following.

* Two E81 animals also mentioned in Fig. 2.

FIG. 1 a, E94 fetal brain showing the location of the cortical plate and germinal zones of presumptive striate cortex as revealed by acetylcholine esterase histochemistry. Arrows indicate the borders of the striate cortex. b, Distribution of the ³H-thymidine labelling in the visual cortex of a newborn monkey (E165) after injection at E81. Small dots represent neurons containing 3–9 grains, open circles represent those containing 10–20 grains and crosses those containing >20 grains. Labelled neurons in striate cortex are nearer the surface than in extrastriate cortex, and are nearly double the number of those in extrastriate cortex. V₁, striate cortex; V₂, extrastriate cortex; 4C, layer 4C of striate cortex; 1, 2, 3, 4 and 5 indicate cortical layers.

METHODS. Cynomolgus monkeys with controlled dates of impregnation were used to provide 7 fetuses of known gestational age (Table 1). Surgery on pregnant monkeys and fetuses was performed as described elsewhere²⁴. Fetuses were injected (i.p.) with ~10–20 μ Ci per g body weight of ³H-thymidine in saline (specific activity, 40–60 μ Ci mmol⁻¹) and replaced in the uterus. In the long-survival group (14–84 days; group I in Table 1) pregnant monkeys were returned to their cages and medicated for two days with an analgesic (Visceralgine, injected i.m.). A muscular relaxant (Duvadilan, also given i.m.) was given twice daily for two weeks and fetuses were either delivered by caesarian section or left until birth (E165). During short survival (E64, 1 h; E78, 7 h; E94, 1.5 h), the pregnant monkey was maintained under anaesthesia. Fetuses were perfused through the heart with 4% paraformaldehyde. In fetuses aged E64, E78, E92 and E94, cells were counted before the appearance of a clear cytoarchitectonic border of striate cortex. Acetylcholine esterase is transiently expressed by extrastriate cortex²⁶ so we used an E94 fetus to determine the borders of presumptive striate cortex in the E64, E78 and E94 fetuses (see panel a). Striate cortex is a large cortical area close to paleocortex, so its position can be extrapolated from E94 to the E64 and E78 fetus. The brain used to detect acetylcholine esterase was cut on a freezing microtome and processed according to ref. 26. Brains used for ³H-thymidine studies were embedded in paraffin wax and 5- or 10- μ m-thick sections were cut. In the long-survival group, sections were processed for autoradiography to give a maximum of 40 grains per differentiated cell. Three intensities of labelling were distinguished: heavy (20–40 grains), moderate (10–19 grains) and light (3–9 grains). ³H-thymidine-labelled cells were counted in 10- or 5- μ m-thick sections counterstained with cresyl violet. This was not possible in the short-survival group because of the small size of the neuroblasts. Sections for quantitative analysis



cut the dorsal cerebral operculum perpendicularly. Observations were made with a time-50 oil objective. Labelled cells were monitored on a video screen and their positions charted by a high-precision plotting system (Biocom, running on HISTO software). Because of local fluctuations in the density of labelled cells in the cortical plate and, to a lesser degree, in the germinal zone, we counted large numbers of labelled cells so as to determine reliably the differences between striate and extrastriate cortex. In the cortical plate, cells were counted in each animal (except the E81 fetus) in 2–7 sections, and in each section in 10–26 100- μ m segments. In the germinal zone, cells were counted in 2 sections per animal and in 4–8 300- μ m wide segments. In the E81 fetus, cells were counted continuously over a large area of striate and extrastriate cortex. Significance levels were determined with a 2-ways ANOVA test, except in the case of E81, for which a Student *t*-test was used. The ratio of labelled cells in equivalent areas of the cortical plate of striate and extrastriate cortex was expressed as a per cent (Fig. 2).

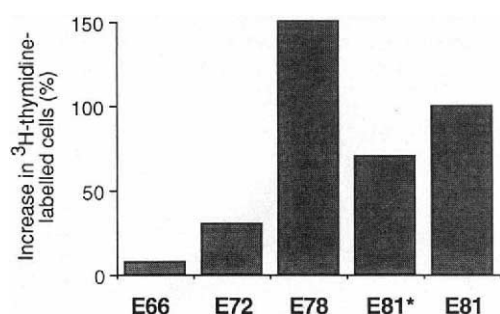


FIG. 2 Histograms showing per cent increase of labelled cells in striate cortex compared with extrastriate cortex after injection at different fetal ages.

when the labelled nucleotide is incorporated into the nucleus⁸⁻¹⁰. After short survival times there was intense labelling in the germinal zones lining the ventricles, and there were more labelled cells in the germinal zone underlying the striate cortex than in that underlying extrastriate cortex (Fig. 3). Although the packing density of precursors in the germinal zone underlying striate cortex was higher than that underlying extrastriate cortex, counts of labelled and unlabelled cells showed that at later stages of corticogenesis the labelling index in the germinal zone underlying striate cortex was significantly higher than that of extrastriate cortex (Table 1). Injection at E78 gave a labelling index in the germinal zone underlying striate cortex that was 49% higher than in the germinal zone underlying extrastriate cortex. Injection at E94 led to a steep drop in the labelling index for both areas, indicating that neurogenesis was slower than in the younger fetus. At this age, the labelling index in the germinal zone of striate cortex was 29% higher than it was in the germinal zone of extrastriate cortex. By contrast, the labelling index at E64 did not differ significantly between the germinal zones of presumptive striate and extrastriate cortex.

The higher labelling index in the germinal zone giving rise to striate cortex indicates an increased proportion of cells in S phase, due either to a shortened cycle time or a maintained cycle duration with a proportionally longer S phase⁸⁻¹⁰. In the later stages of neurogenesis, duration of S phase is considered to remain fairly constant and variations in the cell cycle time to be due mainly to changes in the length of the G1 phase¹¹⁻¹⁴. Hence under such conditions, when the cell cycle is short, S phase takes up a relatively larger part of the cycle and the labelling index is high. The higher rate of cell production characteristic of the germinal zone underlying striate cortex might thus be due to shorter cell-cycle times as well as a higher density of precursors. Differential cell death changes the number of supragranular

neurons in different areas in rodents¹⁵, but this is unlikely to be a factor affecting our results as our fetuses were examined before the principal phase of cell death in the cortex¹⁶.

We have shown that the germinal zone giving rise to striate cortex has a higher rate of cell production than the germinal zone giving rise to the adjacent neocortex. Variations in the labelling index correspond to regional differences in the kinetics of the cell cycle. These differences in neurogenic activity of the germinal zone emerge during the later stages of corticogenesis when the main phase of excess striate neuron production occurs. The specialization of the primate striate area germinal zone ensures that this area has higher numbers of supragranular-layer neurons than other areas³.

Our findings are relevant to theories of the specification of cortex. In rodents, thalamic fibres are important for specifying cortical areas at late stages of development^{17,18}, suggesting that the initially formed cortex is a uniform structure¹, but this view needs to be reconciled with recent findings. Molecular markers have revealed a remarkable degree of early regional specialization of phylogenetic subdivisions of the rodent cerebral cortex^{19,20}. Gross subdivisions of the neocortex also show early specialization²¹. More importantly, early specialization of the immature rodent neocortex distinguishes different areas before the arrival of thalamic afferents²², indicating that areal identity may be at least partially specified before neuroblasts arrive in the cortical plate². A degree of heterogeneity of the germinal layers giving rise to the neocortex has been found²³. We have shown that regional variation of neurogenesis in the germinal zone gives rise to a discrete neocortical area. As thalamic input influences the specification of striate cortex in the monkey^{2,24}, it may regulate neurogenesis in the germinal zone²⁵. □

Received 4 May; accepted 15 September 1993.

- O'Leary, D. D. M. *Trends Neurosci.* **12**, 400-406 (1989).
- Rakic, P. *Science* **241**, 170-176 (1988).
- Rockel, A. J., Hiorns, R. W. & Powell, T. P. S. *Brain* **103**, 221-244 (1980).
- Bisconte, J. C. & Marty, R. *Expl Brain Res.* **22**, 37-56 (1975).
- Angervine, J. B. & Sidman, R. L. *Nature* **192**, 766-768 (1961).
- Rakic, P. *Expl Brain Res. Suppl.* **1**, 439-450 (1976).
- Rakic, P. *Science* **183**, 425-427 (1974).
- Waechter, R. V. & Jaensch, B. *Brain Res.* **46**, 235-250 (1972).
- Schultze, B. & Korr, H. *Cell Tissue Kinet.* **14**, 309-325 (1981).
- Takahashi, T., Nowakowski, R. S. & Caviness, V. S. J. *Neurosci.* **13**, 820-833 (1993).
- Korr, H. *Adv. Anat. Embryol. Cell Biol.* **61**, 1-70 (1980).
- Laskey, R. A., Fairman, M. P. & Blow, J. J. *Science* **246**, 609-613 (1989).
- Pardee, A. B. *Science* **246**, 603-608 (1989).
- Mac Auley, A., Werb, Z. & Mirkes, P. E. *Development* **117**, 873-883 (1993).
- Finlay, B. L. & Slatyer, M. *Science* **219**, 1349-1351 (1983).
- Williams, R. W., Ryder, K. & Rakic, P. *Soc. Neurosci. Abstr.* **13**, 1044 (1987).
- O'Leary, D. D. M. & Stanfield, B. B. J. *Neurosci.* **9**, 2230-2246 (1989).
- Schlaggar, B. L. & O'Leary, D. D. M. *Science* **252**, 1556-1560 (1991).
- Barbe, M. F. & Levitt, P. J. *Neurosci.* **11**, 519-533 (1991).
- Ferri, R. T. & Levitt, P. *Cereb. Cortex* **3**, 187-198 (1993).
- Arimatsu, Y. et al. *Proc. natn. Acad. Sci. U.S.A.* **89**, 8879-8883 (1992).
- Cohen-Tannoudji, M. thesis, Paris Univ. (1992).
- Johnston, J. G. & Van Der Kooy, D. *Proc. natn. Acad. Sci. U.S.A.* **86**, 1066-1070 (1989).
- Dehay, C., Horsburgh, G., Berland, M., Killackey, H. & Kennedy, H. *Nature* **337**, 265-267 (1989).
- Kennedy, H. & Dehay, C. *Cereb. Cortex* **3**, 171-186 (1993).
- Kostovic, I. & Rakic, P. *J. Neurosci.* **4**, 25-42 (1984).

ACKNOWLEDGEMENTS. This work was supported by grants from the EEC and MRE. We thank G. Clain for animal care, J. L. Borach for photographic assistance and J. Bullier for critically reading the manuscript.

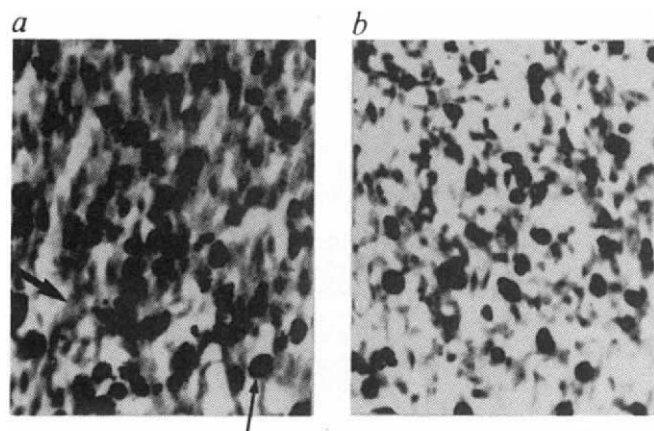


FIG. 3 High-power microphotograph illustrating cell density and ³H-thymidine labelling in the germinal zone of E78 fetus. a, Striate cortex; b, extrastriate cortex. Thin arrow, ³H-thymidine-labelled cell; thick arrow, non-labelled cell. Scale bar, 50 µm.

Neural basis of saccade target selection in frontal eye field during visual search

Jeffrey D. Schall & Doug P. Hanes

Department of Psychology, Wilson Hall, Vanderbilt University, Nashville, Tennessee 37240, USA

CONSPICUOUS visual features commonly attract gaze^{1,2}, but how the brain selects targets for eye movements is not known. We investigated target selection in rhesus monkeys performing a visual search task³ by recording neurons in the frontal eye field, an area known to be responsible for generating purposive eye movements^{4,5}. Neurons with combined visual- and eye movement-related activity were analysed. We found that the initial visual responses to search stimulus arrays were the same whether the target or a distractor was in the response field. We also found that the neural activity evolved to specify target location before the execution of eye movements, ultimately peaking when the target was in the response field and being suppressed when the target was beside but not distant from the response field. These results demonstrate a possible mechanism by which a desired target is fixated and inappropriate eye movements are prevented.

Neurons were monitored in two rhesus monkeys in the rostral bank of the arcuate sulcus where low-intensity microstimulation elicited saccadic eye movements; this property functionally defines the frontal eye field (FEF)⁶. Neurons that were active both in response to visual stimuli and preceding saccadic eye movements were analysed because they were active during target identification and response selection. The activity of neurons was compared during a simple detection task and during a more natural search task (Fig. 1). The activity of one neuron is shown in Fig. 2. In detection trials this cell was active only when the target fell in its response field. In search trials this neuron initially responded to either the target or a distractor in its response field, but the activation changed during the latent period before the movement. If the stimulus in the response field was the target, the discharge rate peaked. If the target was distant from the response field, activation evoked by the distractor in the response field continued through the saccade. If the target appeared near but not centred in the response field, the neuron discharged an abbreviated burst and was then inactive until the saccade.

To quantify the activity associated with visual responses and saccade initiation, discharge rate was measured during a 100-ms interval following stimulus presentation and for 100 ms preceding saccade initiation (Fig. 3). During detection trials neural activity varied with target direction (Fig. 3*a, b*). Differential discharge rates were evident as soon as the cell responded to the target stimulus and continued until the saccade was initiated (Fig. 3*a*). During search trials the spatial and temporal pattern of activation was markedly different. Either the target or a distractor in the response field elicited equivalent activation; thus, for 100 ms following the appearance of the search array, target location could not be identified on the basis of neural activity (Fig. 3*c, d*). Such a lack of significant variation of the initial visual response to the search array was evident in 25 of 28 neurons recorded in two monkeys (ANOVA, $P > 0.05$). But within 100 ms of the saccade, the pattern of neural activity had changed to vary systematically with target location. The presaccadic activation of 27 of the 28 neurons varied significantly with target direction during search trials (ANOVA, $P < 0.05$). Moreover, the pattern of variation of saccade-related activation in search trials was different from that observed in detection trials due to the suppression of the activation elicited by the distractor in the response field when the target was near the response field. This suppression was observed in 22 of the 27 neurons with significant presaccadic variation. The five neurons in which sup-

pression was not observed had unusual response fields that were either unusually large or located in the ipsilateral hemifield. Suppression strength was measured as the ratio of the activation when the target was beside the response field to the activation when the target was in the hemifield opposite the response field, after subtracting baseline activity from both; cells with no suppression had a ratio of 1.0. The mean $\pm$ s.e.m. suppression ratio was 0.66 ± 0.04 (minimum = 0.08).

Visual guidance of eye movements requires information about what is where in the visual field. After each fixation, the target of the next fixation can be selected only as such information becomes available. During visual search, the FEF visual-movement neurons we recorded exhibited a transition from an early nonspecific response to a pattern of activation that specified the target for the upcoming saccade. A similar evolution of neural activity has been observed in inferior temporal cortex during visual search⁷, which may influence FEF neurons. Likewise, the growth of activity of presaccadic movement neurons in the superior colliculus reflects the presence of a target or distractor in the movement field^{8,9}. The FEF heavily innervates the superior colliculus¹⁰, so the activation of colliculus neurons may reflect the response selection we observed in the

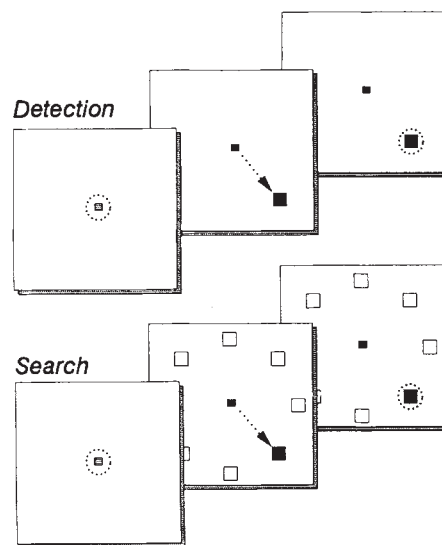


FIG. 1 Detection (top) and search (bottom) trial displays. The dotted circle indicates the required focus of gaze in each interval, and the dotted arrow indicates the saccadic eye movement. Each trial began with presentation of a coloured spot on a video monitor (front panels). After fixation of this spot for 300–500 ms, a target appeared at 1 of 8 locations of equal eccentricity (middle panels). A change of colour of the fixation spot permitted monkeys to shift gaze; for all of the data reported in this paper the colour change of the fixation spot was simultaneous with presentation of the stimuli. Juice reward was given for fixating the target with a single saccadic eye movement with a latency less than 800 ms (rear panels). Baseline neural activity was measured during the fixation period before the target and distractor stimuli were presented. On detection trials the target was presented alone to map the spatial extent of response fields and to identify the type of neuron being recorded. A number of different populations of neurons have been identified in FEF^{4,5}. The focus of this study was visual-movement neurons that began to discharge a consistent interval after stimulus presentation and continued firing until the saccade was generated. On search trials the target was presented with 7 distractors all of the same eccentricity; the stimuli were positioned at the optimum eccentricity for each neuron. Stimuli were scaled in size for eccentricity to yield equivalent performance. During search trials the target was distinguished from distractors by colour (red/green or magenta/cyan) or spatial frequency (low/high). Detection and search trials were run in separate blocks. Standard methods for collecting single-neuron activity and eye position were used¹⁶.

FIG. 2 Neural activity is shown during detection (left) and search (right) trials with the illustrated stimulus configurations. The target was a green 1° square presented either alone or with 7 red 1° square distractors at 10° eccentricity. In the raster displays vertical tickmarks represent times of neuronal discharges, one row per trial. Saccade initiation is indicated by a circle. The rasters are aligned on target presentation and sorted according to saccade latency. Superimposed on the raster is the average spike density function obtained by convolving each spike train with a gaussian filter ($\sigma = 10$ ms). The neuron responded before saccades to a lone target only when it fell in the lower right quadrant (upper left panel); the neuron began to fire with a consistent latency of ~ 90 ms after the target appeared and continued firing until the saccade was executed. Activity remained at baseline when the target appeared at other locations (middle and lower left panels). During search trials a stimulus appeared in the cell's response field on every trial. When the target was in the response field (top right panel), the activation was similar to that observed in the detection trials with the target at the same location (top left panel). When the target was distant from the response field (bottom right panel), the neuron responded to the distractor in the response field and continued discharging at a reduced rate until the saccade. But when the target fell beside the response field (middle right panel), the neuron discharged only briefly, becoming inactive before the saccade.

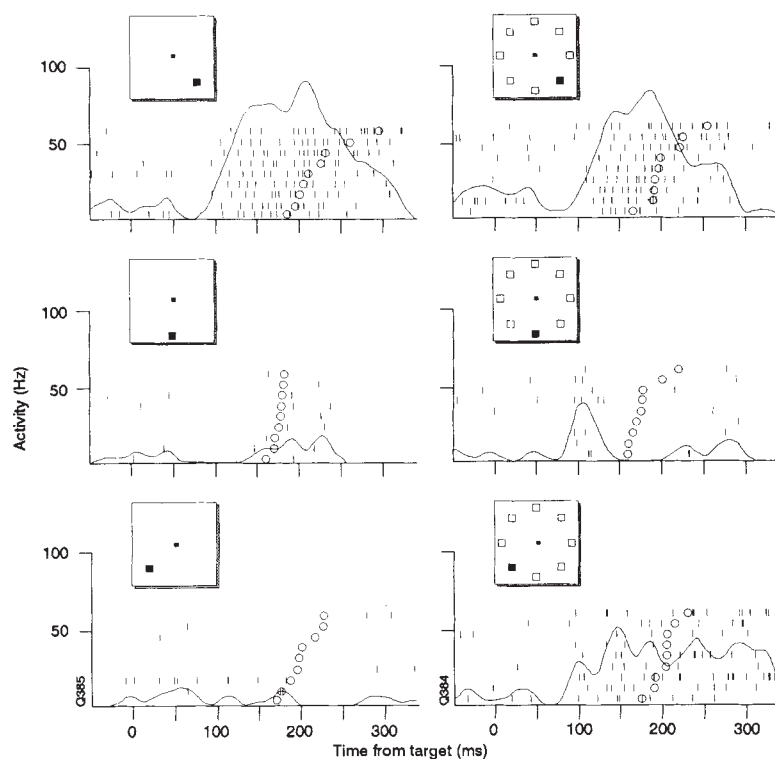
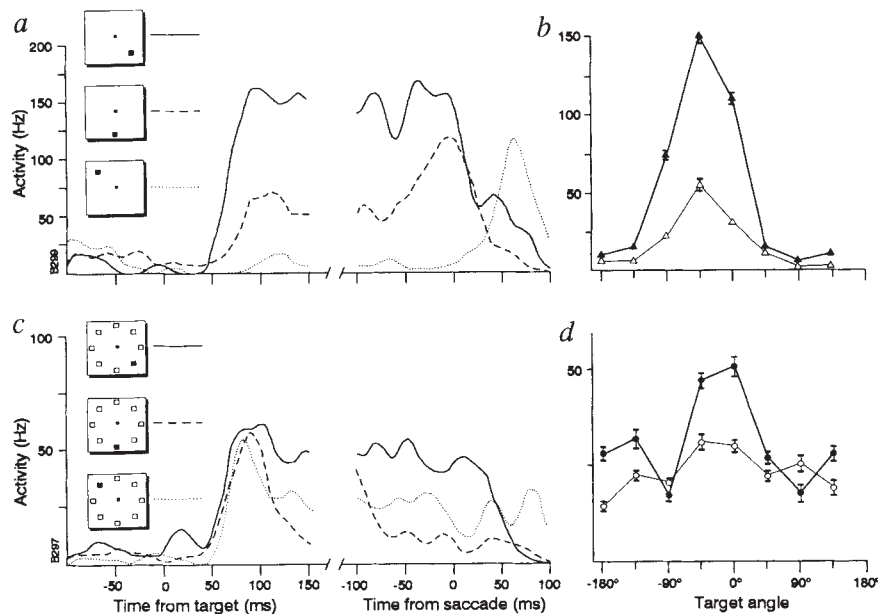


FIG. 3 Evolution of target selection in another FEF neuron. The target was an achromatic 0.33° square, 3 cycle deg^{-1} checkerboard presented alone or with 7 0.33° square, 4.5 cycle deg^{-1} distractors at 4° eccentricity. Average spike density functions are shown for detection trials (a) and search trials (c) aligned on the time of target presentation (left) or saccade initiation (right). Different line types show the activation associated with the three illustrated stimulus configurations. During detection trials both the visual and the saccade-related activity decreased as the target was presented further from the neuron's response field which was located in the lower right quadrant. During search trials an initial visually evoked response occurred in all trials, and the magnitude of this early activation did not vary with search stimulus configuration. But in the interval before the saccade was initiated the activation of the neuron changed according to the location of the target relative to the response field, being maximal when the target was in the response field (solid line), and suppressed when the target was beside (dashed line) but not distant from (dotted line) the response field. Average discharge rate $\pm$ s.e.m. for 100 ms following stimulus presentation (open symbols) and average discharge rate $\pm$ s.e.m. for 100 ms before saccade (closed symbols) is plotted as a function of target direction for detection (b) and search (d) trials. 0° is right, and 90° is up. During detection trials the variation of the visually evoked response with target angle was significant (ANOVA $F(7,65) = 16.4$, $P < 0.001$) as was the variation of the saccade-related activation ($F(7,65) = 86.0$, $P < 0.001$). The pattern of variation with target direction resembled a gaussian curve. During search trials, the neuron responded to the stimulus in its



response field on every trial, and the initial response to the stimulus array did not vary significantly with target direction ($F(7,130) = 2.0$, $P > 0.05$). In contrast, the presaccadic activation varied significantly with target location ($F(7,130) = 8.08$, $P < 0.001$). Moreover, the pattern of variation of saccade-related activity with target direction in search trials was markedly different from that in detection trials, resembling a difference-of-gaussian or Mexican hat curve because of the suppression when the target was beside the response field.

FEF. Visual-movement cells, however, probably do not project directly to the superior colliculus or brainstem^{11,12} and thus may represent an intermediate processing stage.

Before saccades executed during visual search, FEF visual-movement neuron activity became suppressed if a salient target was near its response field. Perhaps such suppression reduces the probability of an eye movement to a distracting stimulus within a neuron's response field. The central facilitation and surrounding suppression of visual-movement neuron activ-

ity resembles the centre-surround receptive field organization of neurons in the early visual system¹³. By analogy to these systems, we surmise that the suppression observed in the FEF arises through lateral inhibition. In fact, models of visual search use topographic representations of the locations of conspicuous features in the image derived from lateral inhibition within feature maps^{14,15}. Our findings provide neurophysiological evidence for a mechanism by which conspicuous stimuli in complex scenes may attract gaze. □

Received 12 April; accepted 22 September 1993.

1. Yarbus, A. L. *Eye Movements and Vision* (Plenum, New York, 1967).
2. Engel, F. *Vision Res.* **17**, 95-108 (1977).
3. Treisman, A. Q. J. *exp. Psychol.* **40A**, 201-237 (1988).
4. Bruce, C. J. & Goldberg, M. E. *J. Neurophysiol.* **53**, 603-635 (1985).
5. Schall, J. D. *J. Neurophysiol.* **66**, 559-579 (1991).
6. Bruce, C. J., Goldberg, M. E., Bushnell, C. & Stanton, G. B. *J. Neurophysiol.* **54**, 714-734 (1985).
7. Chelazzi, L., Miller, E. K., Duncan, J. & Desimone, R. *Nature* **363**, 345-347 (1993).
8. Ottes, F. P., Van Gisbergen, J. A. M. & Eggermont, J. J. *Expl Brain Res.* **66**, 465-478 (1987).
9. Glimcher, P. W. & Sparks, D. L. *Nature* **355**, 542-545 (1992).
10. Leichnetz, G. R. & Goldberg, M. E. in *Neuroanatomy of the Oculomotor System* (ed. Büttner-Ennever, J. A.) 365-429 (Elsevier, New York, 1988).

11. Segraves, M. & Goldberg, M. E. *J. Neurophysiol.* **58**, 1387-1419 (1987).
12. Segraves, M. A. *J. Neurophysiol.* **68**, 1967-1985 (1992).
13. Allman, J., Miezen, F. & McGuinness, E. A. *Rev. Neurosci.* **8**, 407-430 (1985).
14. Koch, C. & Ullman, S. *Hum. Neurobiol.* **4**, 219-227 (1985).
15. Cave, K. R. & Wolfe, J. M. *Cognitive Psychol.* **22**, 225-271 (1990).
16. Schall, J. D. *J. Neurophysiol.* **66**, 530-558 (1991).

ACKNOWLEDGEMENTS. We thank D. King for technical assistance, F. Tu for participating in the experiments and R. Blake, K. Cave, M. Goldberg, T. Preuss and K. Thompson for comments on the manuscript. Supported by the National Eye Institute, the Alfred P. Sloan Foundation and the McDonnell-Pew Program in Cognitive Neuroscience.

Transgenic plants expressing a functional single-chain Fv antibody are specifically protected from virus attack

Paraskevi Tavladoraki*, Eugenio Benvenuto*†, Stefania Trinca*, Domenico De Martinis*, Antonino Cattaneo† & Patrizia Galeffi*

* ENEA, Dipartimento Ricerche e Sviluppo Agroindustriali, Divisione Ingegneria Genetica, CP 2400, 00100 Rome, Italy
† SISSA (Scuola Internazionale di Studi Superiori Avanzati), Via Beirut 2/4, 34013 Trieste, Italy

EXPRESSION of viral genes in transgenic plants is a very effective tool for attenuating plant viral infection¹⁻³. Nevertheless, the lack of generality and risk issues related to the expression of viral genes in plants⁴ might limit the exploitation of this strategy. Expression in plants of antibodies against essential viral proteins could provide an alternative approach to engineer viral resistance. Recently, expression of complete⁵⁻⁷ or engineered⁷⁻⁹ antibodies has been successfully achieved in plants. The engineered single-chain Fv antibody scFv (refs 10, 11) is particularly suitable for expression in plants because of its small size and the lack of assembly requirements. Here we present evidence that constitutive expression in transgenic plants of a scFv antibody, directed against the plant icosahedral tobamovirus artichoke mottled crinkle virus, causes reduction of infection incidence and delay in symptom development.

A panel of monoclonal antibodies was raised against the artichoke mottled crinkle virus (AMCV) virion using standard methods¹², and the binding properties of individual antibodies were analysed by enzyme-linked immunosorbent assay (ELISA). One of these, the F8 antibody (belonging to the IgG2b subclass), had the highest affinity for AMCV coat protein in both dissociated and polymerized forms. In addition, epitope mapping experiments showed that the F8 antibody recognizes a highly conserved site on the coat protein that is involved in divalent-cation regulated swelling of tobamovirus¹³. For this study, existing data on the crystal structure¹⁴ of the type-member of the tobamovirus group, the tomato bushy stunt virus (TBSV), and on the

nucleotide sequence of AMCV coat protein¹⁵ were fundamental.

The hybridoma cell line expressing the F8 antibody was selected to isolate full-length complementary DNA clones of the heavy and light immunoglobulin chains. The variable domains (VH and VL) were amplified by polymerase chain reaction (PCR) using 'universal primers' and inserted into vectors for *Escherichia coli* expression of scFv antibody, in which the two variable domains are connected by a linker peptide (Fig. 1).

Bacterial periplasmic extracts were tested for correct expression of the engineered antibody (scFv[F8]) by western blotting (not shown) using the anti-Tag 9E10 antibody (see legend to Fig. 1). The binding activity of the bacterially produced scFv[F8] antibody against AMCV particles was assayed by ELISA. No binding was detected to four other purified proteins (bovine serum albumin, lysozyme, keyhole-limpet haemocyanin, concanavalin A) or to unrelated (cucumber mosaic virus, CMV) and related (TBSV) icosahedral plant viruses (not shown). Thus, scFv[F8] retains the specificity of the parental antibody. Furthermore, a saturation curve radioimmunoassay was used to determine the affinity range of both scFv[F8] and parent antibody. In these assays the amount of bound antibody on antigen adsorbed on solid phase was detected using the recombinant fusion protein LG (ref. 16) as a second radiolabelled reagent. The affinity constant was evaluated as the reciprocal of the antibody concentration giving 50% saturation of binding. Although avidity effects cannot be excluded, the affinity of the scFv was 100-fold lower (10^7 M^{-1}) than that of the original antibody (10^9 M^{-1}).

To determine whether the engineered antibody was effective in neutralizing the virus in the plant, the scFv[F8] construct was cloned into a plant expression vector to yield the pBG-scFv[F8]-BIN plasmid (Fig. 1). This encodes a cytoplasmic version of the scFv[F8] protein, in which the leader sequence for secretion has been deleted. Transgenic *Nicotiana benthamiana* plants, symptomatic hosts for AMCV, were regenerated after *Agrobacterium*-mediated transformation, essentially as described¹⁷. Northern blot analysis revealed that several (40% of the analysed plants) kanamycin-resistant plants accumulated varying levels of scFv[F8] messenger RNA of the predicted size (Fig. 2a). This was confirmed at the protein level by western blotting (maximum expression being 0.1% of total soluble leaf proteins) (Fig. 2b); ELISA provided evidence for binding specificity of the plant-derived scFv[F8] antibody (not shown). Comparison of data from ELISA and western blots demonstrated that all of the expressed antibody fragment in transgenic plants is active.

To verify the effects of the scFv[F8] antibody transgene on

† To whom correspondence should be addressed.

FIG. 1 Constructs for the expression of scFv in bacteria and plants. H, *Hind*III, P, *Pst*I, Bs, *Bst*EII, S, *Sac*I, Xh, *Xho*I, E, *Eco*RI restriction sites.

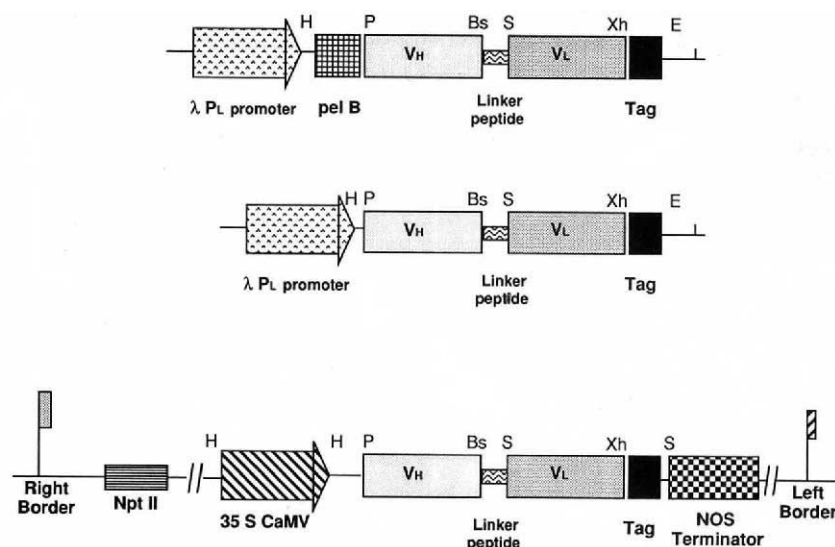
METHODS. Total RNA was extracted from 1×10^8 hybridoma cells using RNagents kit (Promega), and poly(A)⁺-mRNA was isolated through oligo(dT)-cellulose. After cDNA synthesis (cDNA synthesis system plus, Amersham), a plasmid cDNA library was constructed. Heavy and light immunoglobulin chain cDNAs were screened by colony hybridization. Full-length cDNAs were

isolated and sequenced (Sequenase version 2.0, USB). The coding sequences for V_H and V_L have been amplified by PCR according to ref. 21, using as primers: V_H1 FOR 5'-d; (TGAGGAGACGGTGACCGTGGTCCCTTGGCCCCAG), V_H1 BACK 5'-d(AGGTS-MARCTGCAGSAGTCWGG), in which S=C or G; M=A or C; R=A or G; W=A or T, V_k1 FOR 5'-d(CCGTTTGATCTCGAGCTTGGTGCC), V_k1 BACK 5'-d(GACATCGAGCTCACTCAGTCTCCA). The amplified V_H was digested by *Pst*I and *Bst*EII, and the V_L by *Xho*I and *Sac*I. Both fragments were purified through agarose gel electrophoresis and GeneClean kit (Bio 101). Subcloning was done into the pJM1scFv vector (from G. Winter, MRC, Cambridge) for *E. coli* expression, under the control of λ P_L promoter²², which either included or did not include the sequence encoding the pelB signal peptide for secretion²³, yielding the plasmids pJM1scFv[F8](secr) and pJM1scFv[F8](intra), respectively. These vectors contain also the human c-myc Tag peptide²⁴, recognized by the 9E10 monoclonal antibody, and a linker peptide (Gly₂Ser)₃ to connect the V_H carboxy and V_L amino termini. The pJM1scFv[F8](secr) plasmid was

***E. coli*
expression vector**
pJM1-scFv[F8](secr)

pJM1-scFv[F8](intra)

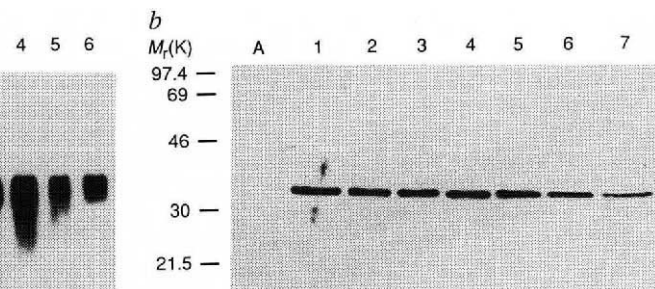
plant expression vector
pBG-scFv[F8]-BIN



introduced into the EMG578 strain of *E. coli*, which expresses a thermo-labile λ repressor. Expression of scFv[F8] antibody in the bacterial periplasm was induced by heating the bacterial culture for 1h at 42 °C. Osmotic shock of the bacterial pellet was done according to ref. 25, which allowed isolation of the soluble periplasmic fraction. Periplasmic extract (20 μ l) was tested for expression of scFv[F8] by western blot analysis (see Fig. 2 legend), whereas 100 μ l of the same extract were tested for AMCV binding activity by ELISA (see Fig. 2 legend). For expression in plants, the scFv[F8] (intra) cassette (fragment *Hind*III-*Eco*RI) lacking the pelB signal sequence was transferred from the pJM1scFv[F8](intra) plasmid to the pBG-dAb vector⁸, yielding the intermediate pBG-scFv[F8] plasmid. This plasmid was completely digested by *Hind*III, the cohesive termini were 'filled in' and scFv[F8] cassette was isolated after *Sac*I partial digestion. This fragment was then ligated to the pBI 121.1 vector²⁶, digested before by the enzymes *Sma*I and *Sac*I, to yield the binary vector pBG-scFv[F8]-BIN.

FIG. 2 Expression of scFv[F8] gene in transgenic *N. benthamiana* plants. **a**, Northern blot analysis. Lane T1, *in vitro* transcript (0.5 ng); lane T2, *in vitro* transcript (0.1 ng); lane M, molecular mass markers; lane A, untransformed plant; lane 1, transgenic line 67i; lane 2, transgenic line 175i; lane 3, transgenic line 80i; lane 4, transgenic line 15s; lane 5, transgenic line 49s; lane 6, transgenic line 86s. **b**, Western blot analysis. Lane A and lanes 1-6, as above; lane 7, affinity purified scFv[F8] from *E. coli* (0.2 μ g).

METHODS. Total RNA was extracted from transgenic or control plant leaves using the RNagents kit (Promega). Total RNA (10 μ g) was subjected to electrophoresis in a denaturing 1% agarose gel, vacuum-blotted onto a nylon membrane (Hybond N, Amersham) and hybridized using the *Eco*RI-*Hind*III scFv[F8] fragment of the pJM1scFv[F8] plasmid as a probe. The 0.24-9.5-kb RNA ladder (BRL) was used as a molecular mass marker. The 980-base *in vitro* transcript of the scFv cassette was also used as a marker of molecular mass and quantity. *In vitro* transcription was done under the control of the Sp6 promoter following the Promega protocol. Total soluble proteins were extracted from transformed and untransformed plant leaves by homogenization in 50 mM Tris-HCl pH 7.5, 0.5 M NaCl, 0.25% Nonidet P-40, 1.3% polyvinylpyrrolidone (PVP), 1 mM phenylmethylsulphonyl fluoride (PMSF), 3 μ g ml⁻¹ pepstatin, 1 μ g ml⁻¹ leupeptin, 5 mM ascorbic acid, and centrifuged at 20,000 g for 30 min at 4 °C. Total protein concentration was calculated in the supernatant by the Bradford method. Equal amounts of total protein from all samples were incubated with CNBr-activated 4B Sepharose (Pharmacia), coupled to 9E10 monoclonal antibody, overnight at 4 °C by constant rotation. After resin washing, the scFv[F8]



protein was eluted by 200 mM glycine, pH 3.0. The eluate was immediately neutralized to pH 7.5 and $10 \times$ PBS was added to a final concentration of $1 \times$ PBS. An aliquot of the eluted proteins was subjected to SDS-PAGE (12%) and transferred electrophoretically to nitrocellulose membrane (Hybond-C super, Amersham). Apparent M_r of proteins were estimated by Rainbow markers (Amersham). Western blots were pre-incubated overnight at 4 °C with 3% BSA, 0.3% gelatine, 10 μ g ml⁻¹ goat total IgG, $1 \times$ PBS before the addition of antibodies. Blots were then incubated with 9E10 monoclonal antibody followed by a biotinylated sheep anti-mouse whole immunoglobulin antibody (Amersham) and by streptavidin-conjugated horseradish peroxidase (Amersham). Peroxidase activity was determined by ECL detection system (Amersham). The rest of the eluted proteins were added in serial dilutions to ELISA microtitre wells (Dynatech) coated with 3 μ g ml⁻¹ AMCV virus and blocked by 2% low-fat dry milk, 0.1% gelatine in $1 \times$ PBS. After overnight incubation at 4 °C, the 9E10 monoclonal antibody was added. A peroxidase-conjugated goat anti-mouse antibody (KPL) was used as secondary antibody. Bound scFv[F8] was detected according to the manufacturer's instructions (KPL).

AMCV infection, challenge experiments were done on protoplasts and whole plants (propagated *in vitro*) from several of the best producers. Initially, protoplasts isolated from transgenic and untransformed plants were infected by AMCV virions. Protoplasts carrying the scFv[F8] construct accumulated smaller amounts of viral coat protein (Fig. 3a) and showed a lower frequency of infection (Fig. 3b) compared to untransformed control, whereas protoplasts infected with a heterologous virus (CMV) supported viral infection at the same level, in both transgenic and control lines (Fig. 3c). This provides evidence that the drastic reduction of AMCV accumulation is associated with the AMCV-binding specificity of the scFv[F8] antibody and it is not likely to be the result of induction of a nonspecific host defence response. An additional proof of this emerged from experiments in which infected protoplasts of *N. benthamiana* plants expressing an antibody fragment with different specificity⁸ accumulated equivalent amounts of AMCV compared to control (not shown). Experiments were also done on primary transformed plants. Two leaves were inoculated by rubbing them with a suspension containing $4 \mu\text{g ml}^{-1}$ AMCV (1,000-fold higher than the minimum concentration necessary to induce systemic infection in *N. benthamiana*) and symptom development was followed daily. The transgenic lines reproducibly developed symptoms (necrosis on uninoculated apical leaves) from 5 to 14 days later than did any of the control plants over a 30-day period of observation.

To confirm the results obtained on primary (R_0) transgenic plants, infection experiments were continued on seedling progeny (R_1) of self-fertilized transgenic lines. Seedlings were classi-

fied as expressors on the basis of ELISA test (see legend to Fig. 2). Seedlings at the three- to four-leaf stage were inoculated with $0.04 \mu\text{g ml}^{-1}$ AMCV. This inoculum concentration was selected after titration experiments and ensures 100% systemic infection on control plants. Our results demonstrated that R_1 progeny of transgenic plants shows lower virus accumulation (Fig. 4b), a highly reduced incidence of infection and a marked delay in symptom appearance (Fig. 4c) compared to the untransformed plants. The protection observed in R_1 plant generation is as strong as in R_0 plants and could not be overcome by application of more concentrated viral inocula (maximum $4 \mu\text{g ml}^{-1}$). All non-expressing F_1 segregants failed to show protection and became infected as control plants. Finally, we checked transgenic plants for aspecific protection. Transgenic and control plants challenged with CMV ($10 \mu\text{g ml}^{-1}$) developed symptoms simultaneously (not shown).

To our knowledge this is the first demonstration of a plant phenotype with an attenuation of viral infection derived from a constitutively expressed, virus-specific antibody. The virtually unlimited repertoire of antibody specificity could thus be exploited to confer new desired phenotypic traits in transgenic plants. Crucial for the success of this strategy seems to be the utilization of the scFv antibodies, which proved to be functionally stable in the cell cytoplasm, in agreement with previous data⁹. Hence, the intracellular scFvs seem particularly suitable to 'immunomodulate' selected cytoplasmic antigens, unlike whole antibody molecules, which need to be targeted to the endoplasmic reticulum for correct assembly/folding and stable accumulation in plants^{5,7}.

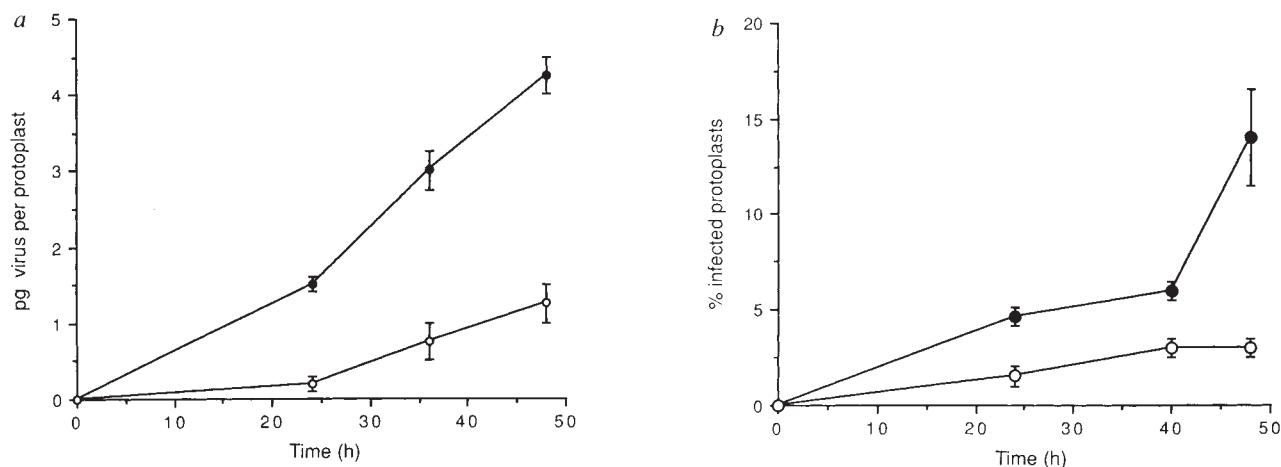


FIG. 3 Relative susceptibility of transgenic *N. benthamiana* protoplasts to infection by AMCV and CMV virus. a, b, Infection of transgenic and untransformed protoplasts by AMCV virus. c, Comparison between AMCV and CMV accumulation in transgenic and untransformed protoplasts. Data relate to plant 67i as a representative of all expressing lines tested. Each point is the average value from three independent experiments. Bars indicate the standard deviation. Untransformed control, filled circle; transgenic line 67i, empty circle.

METHODS. Protoplasts were isolated according to ref. 27 from plant leaves propagated *in vitro* and were infected with $8 \mu\text{g}$ AMCV virus or $1 \mu\text{g}$ CMV virus per 10^6 protoplasts according to ref. 28. Infection of protoplasts was assayed at various time intervals (a, b) or 21 h (c) after virus inoculation. The virus accumulation was determined by indirect double antibody sandwich ELISA according to ref. 29 and expressed as pg accumulated virus per live protoplast (a, c). The percentage of infected protoplasts was determined by immunofluorescence assay³⁰ using the F8 monoclonal antibody followed by a fluorescein-conjugated goat anti-mouse IgG (Pierce) (b). The number of surviving protoplasts was equal within experimental groups.

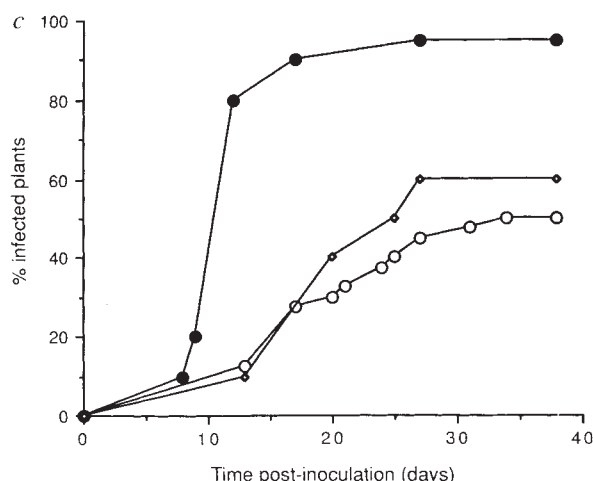
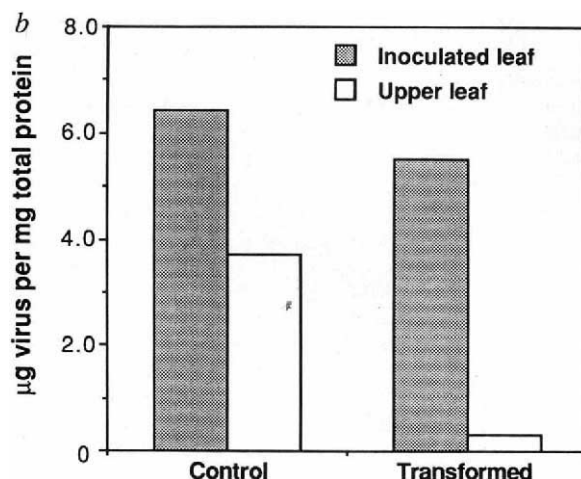


FIG. 4 a, Effect of AMCV infection of *N. benthamiana* plants 14 days after inoculation. Right, untransformed control; left, R_1 transgenic plant expressing the scFv[F8] antibody. b, AMCV accumulation in inoculated and second upper leaves of infected plants shown in a. c, Percentage of untransformed and R_1 transgenic plants showing AMCV infection at successive days after inoculation with AMCV virions. Untransformed control, filled circles; transgenic line 67i, empty circles; transgenic line 175i, diamonds.

METHODS. Seeds from untransformed and self-fertilized (R_1) transgenic *N. benthamiana* plants were germinated and assayed for scFv[F8] antibody by ELISA (see Fig. 2 legend). Three days later, the two youngest expanded leaves were inoculated using carborundum as an abrasive with a suspension of purified AMCV virions at a concentration of $0.04 \mu\text{g ml}^{-1}$. The plants were then placed in a growth chamber under light-dark cycles of 16 h of light and 8 h of darkness, at 12 W m^{-2} light intensity, 22°C and relative humidity 80%. Plants were scored as having positive symptoms when apical leaves demonstrated necrosis. Various transgenic lines from the best producers were used and two to three independent experiments were done from each line. Each experiment included 25–40 seedlings from every plant line. The results were analysed for significance by the Kaplan–Meier method ($P=0.0001$). For virus quantification, leaves were cut and homogenized as described in Fig. 2 legend; ELISA was done as described for Fig. 3.

The question as to how the intracellular ‘plantibody’ confers resistance remains unanswered. To identify the mechanism(s), it seems important to acquire information on the virus–antibody interaction during early infection events. It is conceivable that the binding of the scFv[F8] antibody to the Ca^{2+} -binding sites on the coat protein abrogates a critical step in the uncoating of the virus¹⁸ or in the assembly of progeny virus.

Furthermore, understanding the involvement of the coat protein in the intracellular and intercellular traffic of the virus as well as determining the precise location of the scFv[F8] protein, will also help in unravelling the mechanism of protection.



Indeed, the compartmentalization of the antibody within the plant may be important in overall protection. We are currently addressing the question of whether the protection obtained by the expression of cytoplasmic antibody domains is altered when the same molecules are targeted to the extracellular environment.

Regardless of the precise mechanism(s) of protection, the approach of ‘engineered intracellular immunization’^{19,20} seems to be of great potential in conferring new plant defences useful for breeding and could be a valuable tool for understanding basic virus–plant interactions. □

Received 18 May; accepted 12 October 1993.

1. Beachy, R. N., Loesch-Fries, S. & Turner, N. E. *A. Rev. Phytopathol.* **28**, 451–474 (1990).
2. Hanley-Bowdoin, L. & Hemenway, C. in *Genetic Engineering with Plant Viruses* (eds Wilson, T. M. A. & Davies, J. W.) 251–295 (CRC Press, Boca Raton, 1992).
3. Wilson, T. M. A. *Proc. nat. Acad. Sci. U.S.A.* **90**, 3134–3141 (1993).
4. de Zoeten, G. A. *Phytopathology* **81**, 585–586 (1991).
5. Hiatt, A., Cafferkey, R. & Bowdish, K. *Nature* **342**, 76–78 (1989).
6. Düring, K., Hippe, S., Kreuzaler, F. & Schell, J. *Pl. molec. Biol.* **15**, 281–293 (1990).
7. De Neve, M. et al. *Transgenic Res.* **2**, 227–237 (1993).
8. Benvenuto, E. et al. *Pl. molec. Biol.* **17**, 865–874 (1991).
9. Owen, M., Gandechea, A., Cockburn, B. & Whitlam, G. *Biotechnology* **10**, 790–794 (1992).
10. Bird, R. E. et al. *Science* **242**, 423–426 (1988).
11. Winter, G. & Milstein, C. *Nature* **349**, 293–299 (1991).
12. Harlow, E. & Lane, D. in *Antibodies: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, New York, 1988).
13. Hogle, J., Kirchhausen, T. & Harrison, S. C. *J. molec. Biol.* **171**, 95–100 (1983).
14. Harrison, S. C., Olson, A. J., Schutt, C. E., Winkler, F. K. & Bricogne, G. *Nature* **276**, 368–373 (1978).
15. Tavazza, M., Luciolli, A., Ancora, G. & Benvenuto, E. *Pl. molec. Biol.* **13**, 685–692 (1989).
16. Kihlberg, B. M., Sjöbring, U., Kastern, W. & Björk, L. *J. biol. Chem.* **267**, 25583–25588 (1992).
17. Tavazza, R., Ordas, R. J., Tavazza, M., Ancora, G. & Benvenuto, E. *J. Pl. Physiol.* **133**, 640–644 (1988).

18. Rossmann, M. G. & Johnson, J. E. *A. Rev. Biochem.* **58**, 533–573 (1989).
19. Biocca, S., Neuberger, M. S. & Cattaneo, A. *EMBO J.* **9**, 101–108 (1990).
20. Werge, T. M., Biocca, S. & Cattaneo, A. *FEBS Lett.* **274**, 193–198 (1990).
21. Ward, E. S., Güssow, D., Griffiths, A. D., Jones, P. T. & Winter, G. *Nature* **341**, 544–546 (1989).
22. Remaut, E., Stanssens, P. & Fiers, W. *Gene* **15**, 81–93 (1981).
23. Lei, S.-P., Lin, H.-C., Wang, S.-S., Callaway, J. & Wilcox, G. J. *Bact.* **169**, 4379–4383 (1987).
24. Munro, S. & Pelham, H. *Cell* **46**, 291–300 (1986).
25. Plückthun, A. & Skerra, A. *Meth. Enzym.* **178**, 497–515 (1989).
26. Jefferson, R. A., Kavanagh, T. A. & Bevan, M. W. *EMBO J.* **6**, 3901–3907 (1987).
27. Zhu, X.-Y. & Negruțiu, I. in *A Laboratory Guide for Cellular and Molecular Plant Biology* (eds Negruțiu, I. & Gharti-Chhetri, G.) 18–27 (Birkhäuser, Basel, Boston and Berlin, 1991).
28. Maule, A. J. *J. gen. Virol.* **64**, 2655–2660 (1983).
29. Sela, I. *Meth. Enzym.* **119**, 734–744 (1986).
30. Otsuki, Y. & Takebe, I. *Virology* **38**, 497–499 (1969).

ACKNOWLEDGEMENTS. We thank many colleagues who helped to make this work possible, in particular: M. Barba, R. Cettomai, R. De Santis, M. Di Loreto, S. Lucetti, C. Mancini, M. Mariani, S. Marini, F. Moretti, D. Pashkoulou, M. Pezzotti, M. Perone-Pacífico and M. Piscitelli and V. Di Rita, R. Koes and C. Mariani for suggestions. This work was partially supported by grants from CNR-RAISA and EEC-Biotechnology. Fellowships were awarded by UNIDO (P.T.) and Fondazione Adriano Buzzati-Traverso (D.D.).

Heregulin induces tyrosine phosphorylation of HER4/p180^{erbB4}

Gregory D. Plowman, Janell M. Green,
Jean-Michel Culouscou, Gary W. Carlton,
Victoria M. Rothwell & Sharon Buckley

Bristol-Myers Squibb Pharmaceutical Research Institute,
3005 First Avenue, Seattle, Washington 98121, USA

THE *HER4/ERBB4* gene encodes a 180K transmembrane protein (HER4/p180^{erbB4}) that is structurally related to the 185K product (HER2/p185^{erbB2}) of the *HER2/ERBB2* proto-oncogene¹. A 45K heparin-binding glycoprotein (p45) has been characterized that specifically activates the intrinsic tyrosine kinase activity of HER4 (ref. 2). This HER4 ligand shares several features with the heregulin family of proteins, including molecular mass, ability to induce differentiation of breast cancer cells, activation of tyrosine phosphorylation in MDA-MB453 cells, and amino-terminal protein sequence. Heregulin exists as multiple isoforms and all are presumed to interact directly with HER2 (refs 3–6). We have used binding and phosphorylation studies with recombinant ligand on cell lines expressing recombinant receptors, and report here that heregulin, like p45, is a specific ligand for HER4. Furthermore, heregulin fails to induce phosphorylation of HER2 in the absence of HER4. These findings suggest that activation of the HER4 receptor is involved in signal transduction by heregulin.

The heregulins are a family of molecules that were first isolated as specific ligands for HER2^{3–6}. A rat homologue was termed Neu differentiation factor because of its ability to induce differentiation of breast cancer cells through its interaction with HER2/Neu³. Heregulin also appears to be important in development and maintenance of the nervous system, based on its abundant expression in cells of neuronal origin and because alternatively spliced forms of the heregulin gene encode two neurotrophic activities. One neural-derived factor is termed acetylcholine receptor inducing activity (ARIA)⁵. This heregulin isoform is responsible for stimulation of neurotransmitter receptor synthesis during formation of the neuromuscular junction. A second factor is called glial growth factor (GGF), reflecting the proliferative effect this molecule has on glial cells in the central and peripheral nervous system⁶. Additional, less well characterized molecules that appear to be isoforms of heregulin include p45, gp30 and p75 (refs 1, 2, 7, 8).

The nature of heregulin's interaction with HER2 has recently been questioned because several HER2-neutralizing antibodies fail to block heregulin activation of human breast cancer cells, and heregulin only activates tyrosine phosphorylation of HER2 in cells of breast, colon and neuronal origin, and not in fibroblasts or ovarian cell lines that overexpress recombinant HER2⁹. These studies suggest that HER2 alone is not sufficient for binding and activation by heregulin.

To determine whether HER4 is involved in signalling by heregulin, we examined the ability of recombinant rat heregulin to stimulate tyrosine phosphorylation in a panel of Chinese hamster ovary (CHO) cells that ectopically express human HER2 or HER4. The activity of recombinant heregulin was first confirmed by its ability to induce tyrosine phosphorylation of a high-molecular-mass protein in MDA-MB453 cells (Fig. 1a). Heregulin had no effect on CHO cells expressing only HER2 (Fig. 1b), but these cells must have a functional receptor because their tyrosine kinase activity could be stimulated by either of two antibodies specific to the extracellular domain of HER2 (Fig. 1b). Heregulin was able to induce tyrosine phosphorylation of a 180K protein in CHO cells expressing HER4 (Fig. 1a).

Species differences in ligand–receptor interactions have been reported for the epidermal growth factor (EGF) receptor¹⁰. It is unlikely that such differences are responsible for our failure to detect a direct interaction between rat heregulin and human HER2, because previous studies have shown that rat heregulin does not directly interact with rat HER2/Neu⁹. In addition, rat, rabbit and human heregulin share high sequence homology and induce tyrosine phosphorylation in their target cells of human origin^{3–5}.

These findings support the earlier observation that HER2 alone is not sufficient to transduce the heregulin signal. To address this possibility further, we established a panel of human CEM cells that express the recombinant receptors either alone or in combination. We used a model system of human origin

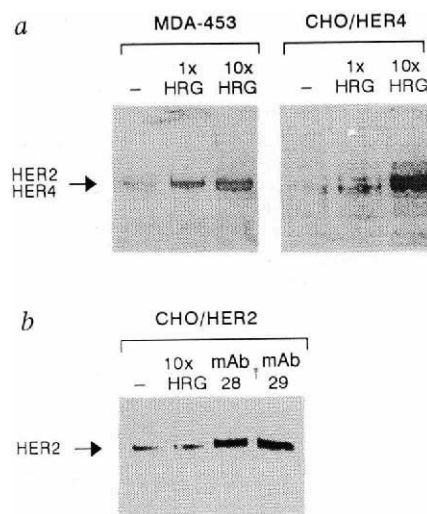
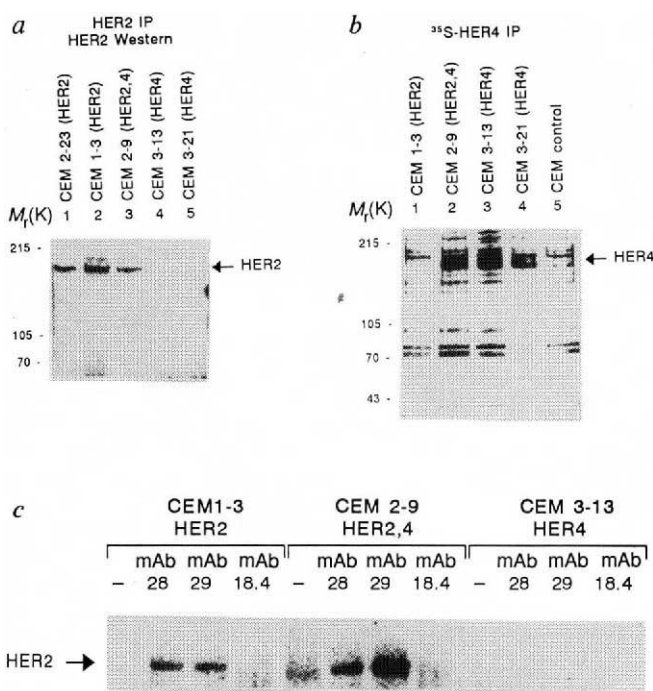


FIG. 1 Recombinant heregulin induces tyrosine phosphorylation of HER4. Tyrosine-phosphorylated receptors were detected by western blotting with an anti-phosphotyrosine monoclonal antibody (mAb). a, Monolayers of MDA-MB453 or CHO/HER4 cells were incubated with media from COS-1 cells transfected with a rat heregulin expression plasmid (HRG), or with a cDM8 vector control (–). The medium was either applied directly (1x) or after concentrating 20-fold (20x; and vector control). Solubilized cells were immunoprecipitated with anti-phosphotyrosine mAb. b, Monolayers of CHO/HER2 cells were incubated as above with transfected COS-1 cell supernatants or with two stimulatory mAbs to HER2 (mAb 28 and 29). Solubilized cells were immunoprecipitated with anti-HER2 mAb. Arrows indicate the HER2 and HER4 proteins.

METHODS. CHO cells expressing recombinant HER4 or HER2 were generated as previously described¹. Cells (1×10^5 of CHO/HER2 and CHO/HER4, and 5×10^5 of MDA-MB453) were seeded in 24-well plates and cultured for 24 h. Cells were starved in serum-free medium for 1–6 h before addition of conditioned medium from transfected COS cells, or 25 $\mu\text{g ml}^{-1}$ HER2-stimulatory mAb (N28 and N29)¹¹. After 10 min treatment at room temperature, cells were solubilized² and immunoprecipitated with 2 μg anti-phosphotyrosine mAb (PY20; ICN Biochemicals) or anti-HER2 mAb (c-neu Ab-2; Oncogene Sciences) and anti-mouse IgG-agarose (Sigma). Western blots were probed with PY20 as previously described¹, and bands were detected on a Molecular Dynamics phosphorimager. Recombinant rat heregulin was produced as follows. A 1.6-kilobase (kb) fragment encoding the entire open reading frame of rat heregulin (and 324 base pairs (bp) of 5'-untranslated sequence) was obtained by PCR using normal rat kidney RNA as a template. This fragment was inserted into a cDM8-based expression vector (Invitrogen) to generate cNDF1.6. The expression plasmid was introduced into COS-1 cells using the DEAE-dextran/chloroquine method¹². Cells were grown for 2 days in Dulbecco's Modified Eagle Medium (DMEM)/10% FBS, followed by an additional 2 days in serum-free DMEM. Clarified conditioned medium was then dialysed against 0.1 M acetic acid, dried, and resuspended as a 20-fold concentrate in DMEM.

FIG. 2 Expression of recombinant HER2 and HER4 in human CEM cells. Transfected CEM cells were selected that stably express either HER2, HER4 or both recombinant receptors. *a*, Recombinant HER2 was detected by immunoprecipitation of cell lysates with anti-HER2 mAb (Ab-2) and western blotting with another anti-HER2 mAb (Ab-3). *b*, Recombinant HER4 was detected by immunoprecipitation of ^{35}S -labelled cell lysates with HER4-specific rabbit anti-peptide antisera. *c*, Three CEM cell lines were selected that express one or both recombinant receptors and aliquots of each were incubated with media control (-), with two HER2-stimulatory mAbs (mAb 28 and 29), or with an isotype matched control mAb (18.4). Solubilized cells were immunoprecipitated with anti-HER2 mAb (Ab-2) and tyrosine-phosphorylated HER2 was detected by western blotting with an anti-phosphotyrosine mAb. Arrows indicate the HER2 and HER4 proteins.

METHODS. cHER2 and cHER4 expression plasmids were generated by insertion of the complete coding sequences of human HER2 and HER4 into cNEO, an expression vector that contains an SV2-NEO expression unit inserted at a unique *Bam*HI site of cDM8. These constructs were linearized and transfected into CEM cells by electroporation with a Bio-Rad Gene Pulser apparatus essentially as previously described³. Stable clones were selected in RPMI, 10% FBS supplemented with 500 $\mu\text{g ml}^{-1}$ active Geneticin. HER2 immunoprecipitations were done as described for Fig. 1, using 5×10^6 cells per reaction, and the HER2 western blots were probed with a second anti-HER2 mAb (c-neu Ab-3; Oncogene Sciences). For metabolic labelling of HER4, 5×10^6 cells were incubated for 4–6 h in methionine- and cysteine-free minimal essential medium (MEM) supplemented with 2% FBS and 250 $\mu\text{Ci ml}^{-1}$ [^{35}S]Express protein-labelling mix (New England Nuclear). Cells were washed twice in RPMI and solubilized as described. Lysates were then incubated for 6 h, 4 °C with 3 μl each of two rabbit antisera raised against synthetic peptides corresponding to two regions of the cytoplasmic domain of human HER4 (⁸⁶⁴LARLEGDEKEYNADGG⁸⁸⁰ and ^{1,010}EEDLEDMMDAEEY^{1,022}; single-letter amino-acid code). Immune complexes were precipitated with 5 μg goat anti-rabbit immunoglobulin (Cappel) and protein G Sepharose (Pharmacia). Proteins were resolved on 7% SDS-polyacrylamide gels and exposed on the phosphorimager.



For mAb-stimulation assays, 5×10^6 cells were resuspended in 100 μl RPMI and 25 $\mu\text{g ml}^{-1}$ mAb was added for 15 min at room temperature. Control mAb 18.4 was a murine IgG₁ specific to human amphiregulin¹³. After mAb treatment, cells were washed in PBS, solubilized², and immunoprecipitated with anti-HER2 mAb (Ab-2). Tyrosine-phosphorylated HER2 was detected by PY20 western blot as in Fig. 1.

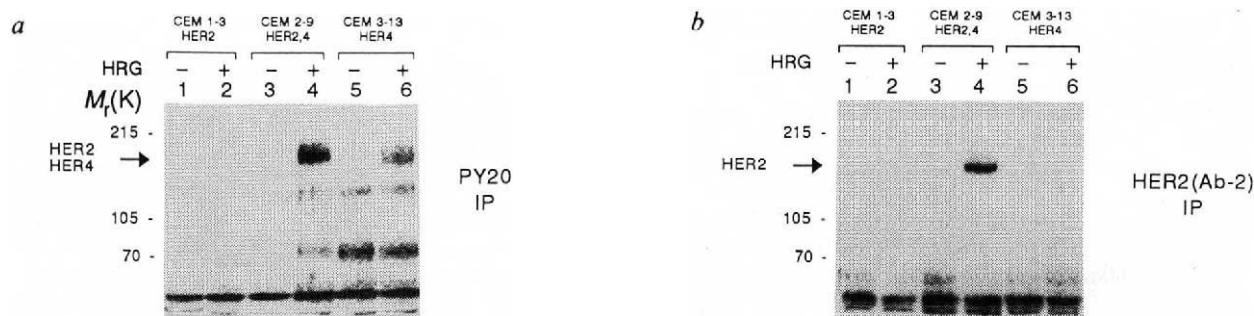
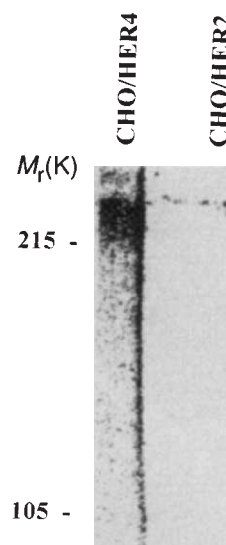


FIG. 3 Heregulin induces tyrosine phosphorylation in CEM cells expressing HER4. Three CEM cell lines that express either HER2 or HER4 alone (CEM 1–3 and CEM 3–13) or together (CEM 2–9) were incubated with 7 $\times$ concentrated supernatants from mock- (–) or heregulin-transfected (+) COS-1 cells. Solubilized cells were immunoprecipitated (IP) with *a*, anti-phosphotyrosine mAb (PY20); *b*, HER2-specific anti-HER2 mAb (Ab-2); or *c*, HER4-specific mAb (6–4). In each case, tyrosine-phosphorylated receptors were detected by western blotting with PY20. Arrows indicate the HER2 and HER4 proteins. HRG, recombinant rat heregulin.

METHODS. Recombinant rat heregulin was prepared as for Fig. 1, and diluted to 7 $\times$ in RPMI. The HER4-specific mAb was prepared by immunization of mice with baculovirus-expressed HER4 (manuscript in preparation). CEM cells (5×10^6) were treated with the concentrated supernatants for 10 min at room temperature and precipitated with PY20 or anti-HER2 mAb (Ab-2) as described in Fig. 1 legend. Immunoprecipitation with anti-HER4 mAb was done by incubation of cell lysates with a 1:5 dilution of hybridoma supernatant for several hours followed by 2 μg rabbit anti-mouse immunoglobulin (Cappel) and protein A-Sepharose CL-4B (Pharmacia). Western blots were probed with PY20 as described for Fig. 1.

FIG. 4 Covalent crosslinking of iodinated heregulin to HER4. ^{125}I -heregulin was added to CHO/HER4 or CHO/HER2 cells for 2 h at 4 °C. Washed cells were crosslinked with bis(sulphosuccinimidyl)-suberate³, lysed, and the proteins separated using 7% PAGE. Labelled bands were detected on the phosphorimager. M_r markers are shown on the left. METHODS. To facilitate purification, recombinant heregulin was produced as an epitope-tagged fusion with amphiregulin. The 63-amino-acid EGF-structural motif of rat heregulin³ from serine 177 to tyrosine 239 was fused to the N-terminal 141 amino acids of the human amphiregulin precursor¹³. This truncated portion of heregulin is active when expressed in *Escherichia coli*⁴, and the N-terminal residues of amphiregulin provide an epitope for immunological detection and purification of the recombinant protein. This cDNA fragment was spliced into a cDM8-based expression vector for transient expression in COS-1 cells. Recombinant heregulin was purified by cation exchange and reversed-phase chromatography and shown to be active by the specific stimulation of HER4 tyrosine phosphorylation. Purified heregulin was iodinated with 250 μCi ^{125}I -labelled Bolton-Hunter reagent (NEN). CHO/HER4 or CHO/HER2 cells were incubated with ^{125}I -heregulin (10^5 c.p.m.) for 2 h at 4 °C. Monolayers were washed in PBS, and 3 mM bis(sulphosuccinimidyl)-suberate (Pierce) was added for 30 min on ice. The cells were washed in Tris-buffered saline, dissolved in SDS sample buffer, run on a 7% polyacrylamide gel, and visualized on the phosphorimager.



because many of the reagents against ErbB family members are specific to the human homologues. CEM cells are a human T-lymphoblastoid cell line which does not express EGF receptor, HER2, HER3 or HER4. Figure 2 demonstrates the selection of three CEM cell lines that express only HER2 (CEM 1–3), only HER4 (CEM 3–13) or both HER2 and HER4 (CEM 2–9). The presence of a functionally and structurally intact HER2 in the appropriate cells was confirmed by the induction of HER2 tyrosine phosphorylation by each of the two antibodies specific for the extracellular domain of HER2, but not by an isotype-matched control antibody (Fig. 2c).

The panel of CEM cells were then analysed by phosphotyrosine western blots of cell lysates after treatment with heregulin and immunoprecipitation with three different monoclonal antibodies. Precipitation with an anti-phosphotyrosine antibody (PY20) again demonstrates that heregulin can stimulate tyrosine phosphorylation in cells expressing HER4, but not in cells expressing only HER2 (Fig. 3a). Precipitation with an antibody specific for the extracellular domain of HER2 demonstrates that HER2 is tyrosine-phosphorylated in response to heregulin in cells that coexpress HER4 (Fig. 3b). Furthermore, precipitation with a HER4-specific antibody confirms that heregulin induces tyrosine phosphorylation of HER4 irrespective of HER2 expression (Fig. 3c). Because of coexpression of HER2 and HER4 in many breast carcinomas¹, these findings suggest that earlier studies of heregulin-HER2 interactions may require re-evaluation.

Previous binding and covalent crosslinking studies have demonstrated that p45 binds specifically to HER4 and displays a single high-affinity site with a dissociation constant (K_d) of 5 nM on CHO/HER4 cells². Preliminary crosslinking studies have been done on these cells with recombinant heregulin revealing a high-molecular-mass species that corresponds to the heregulin-

HER4 receptor complex (Fig. 4).

We have shown here that heregulin induces tyrosine phosphorylation of HER4 in the absence of HER2. In contrast, heregulin does not directly stimulate HER2. However, in the presence of HER4, heregulin induces phosphorylation of HER2, presumably either by transphosphorylation or through receptor heterodimerization. Together, these experiments suggest that HER4 is the receptor for heregulin.

Most breast cancer cells that overexpress HER2 are responsive to heregulin, whereas HER2-positive ovarian and fibroblast lines do not respond to the ligand. This observation could be explained by the fact that HER4 is coexpressed with HER2 in most or all of the breast cancer cell lines studied, but not in the ovarian carcinomas. Furthermore, overexpression of HER2 in heregulin-responsive breast cancer cells leads to increased binding, whereas expression of HER2 in heregulin-unresponsive ovarian or fibroblast cells has no effect⁹. Vital questions remain: does HER2 overexpression force heterodimer formation and increase the number or affinity of heregulin binding sites; is HER4 solely responsible for heregulin-mediated phosphorylation of HER2; does HER4 interact with the EGF receptor or HER3; and is there a specific ligand for HER2?

It will also be crucial to determine which of the many heregulin isoforms interact with HER4. Northern and *in situ* analyses localizes HER4 to the white matter and glial cells of the central and peripheral nervous system, as well as to cardiac, skeletal and smooth muscle¹. This distribution is consistent with HER4 being involved in signalling by the neurotrophic factors GGF and ARIA. Recognition of HER4 as a primary component of the heregulin signal transduction pathway will assist in deciphering the molecular mechanisms that result in its diverse biological effects. □

Received 14 July; accepted 14 October 1993.

1. Plowman, G. D. et al. *Proc. natn. Acad. Sci. U.S.A.* **90**, 1746–1750 (1993).
2. Culouscou, J.-M. et al. *J. Biol. Chem.* **268**, 18407–18410 (1993).
3. Wen, D. et al. *Cell* **69**, 559–572 (1992).
4. Holmes, W. E. et al. *Science* **256**, 1205–1210 (1992).
5. Falls, D. L. et al. *Cell* **72**, 801–815 (1993).
6. Marchionni, M. A. et al. *Nature* **362**, 312–318 (1993).
7. Lupu, R. et al. *Science* **249**, 1552–1555 (1990).
8. Lupu, R. et al. *Proc. natn. Acad. Sci. U.S.A.* **89**, 2287–2291 (1992).
9. Peles, E. et al. *EMBO J.* **12**, 961–971 (1993).

10. Lax, I. et al. *Molec. cell. Biol.* **8**, 1970–1978 (1988).
11. Stancovski, I. et al. *Proc. natn. Acad. Sci. U.S.A.* **88**, 8691–8695 (1991).
12. Seed, B. & Aruffo, A. *Proc. natn. Acad. Sci. U.S.A.* **84**, 3365–3369 (1987).
13. Plowman, G. D. et al. *Molec. cell. Biol.* **10**, 1969–1981 (1990).

ACKNOWLEDGEMENTS. We thank Y. Yarden for anti-HER2 antibodies N28 and N29, J. Blake for peptide synthesis, V. McDonald for production of rabbit antisera, U. Garrigues and S. Jackson for antibody production, M. Neubauer and J. Cook for sequence verification of expression constructs and T. Youngman for oligonucleotide synthesis.

Functional ecdysone receptor is the product of *EcR* and *Ultraspiracle* genes

Tso-Pang Yao^{*†}, Barry Marc Forman, Zeyu Jiang[‡], Lucy Cherbas[‡], J.-Don Chen^{*}, Michael McKeown, Peter Cherbas[‡] & Ronald M. Evans^{*}

^{*} Howard Hughes Medical Institute, The Salk Institute for Biological Studies, La Jolla, California 92037, USA

[†] Graduate Program in Biomedical Sciences, University of California, San Diego, La Jolla, California 92037, USA

[‡] Department of Biology, Indiana University, Bloomington, Indiana 47405, USA

ALTHOUGH the biological activity of the insect moulting hormone ecdysone, is manifested through a hormonally regulated transcriptional cascade associated with chromosomal puffing¹⁻³, a direct association of the receptor with the puff has yet to be established. The cloned ecdysone receptor⁴ (*EcR*) is by itself incapable of high-affinity DNA binding or transcriptional activation. Rather, these activities are dependent on heterodimer formation with *Ultraspiracle*⁵ (*USP*) the insect homologue of vertebrate retinoid X receptor⁶. Here we report that native *EcR* and *USP* are co-



FIG. 1 Differential DNA binding and ligand-dependent transactivation by *EcR* (*GEcR*)/*USP* and *EcR* (*GEcR*)/*RXR*. *a*, *EcR*/*USP* but not *EcR*/*RXR-α* binds to Hsp27-*EcRE*. *In vitro*-translated *EcR* (B1 isoform¹⁹) was incubated with ³²P-labelled Hsp27-*EcRE* with unprogrammed lysate, or with *USP* or human *RXR-α* and analysed by mobility shift assay. Note that only *EcR*/*USP* (lane 3) but not *EcR*/*RXR-α* (lane 4) gives rise to a DNA-binding complex. *b*, Cotransfection of *GEcR* and *RXR-α* does not render CV1 cells responsive to 20E. CV1 cells were transfected with expression vectors for *GEcR* (RSV-*GEcR*), an *EcR* chimera that contains a strong mammalian transactivation domain from glucocorticoid receptor fused to the amino terminus of *EcR*²⁰, either alone or together with *USP* (CMX-*USP*) or h*RXR-α* (CMX-h*RXR-α*). Cotransfection of *GEcR* and *USP* typically gives ~5–8-fold induction of reporter ΔMTV-Hsp27-*EcRE*₅-CAT, whereas cotransfection of *GEcR* and *RXR-α* does not show any induction in response to 20E (20 μM). We note that an *RXR* chimera whose ligand-binding domain is replaced by that of *USP* (XXU⁶) can respond to 20E but not the reverse chimera UUX⁶ (data not shown), indicating that the C terminal/ligand-binding domain of *USP* is required to establish 20E responsiveness. *c*, Muristerone A (MurA) can activate both *GEcR*/*RXR*s and *GEcR*/*USP* heterodimers in CV1 cells. Transfections were performed as described in *b* except that MurA (1 μM; Zamboni, Italy) was included as ligand. Both *GEcR* (this figure) and *EcR* (data

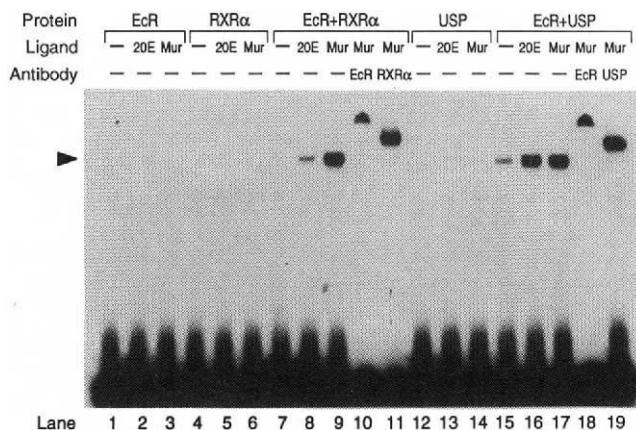
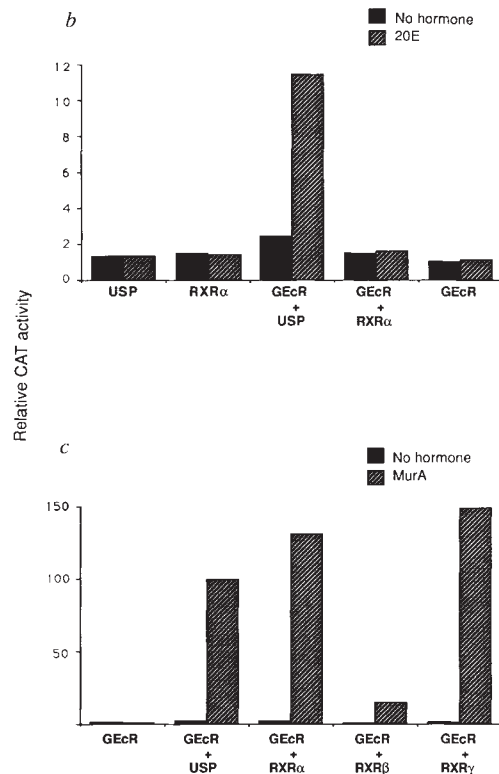


FIG. 2 MurA and 20E differentially stimulate the DNA-binding activity of *EcR*/*USP* and *EcR*/*RXR-α*. *In vitro*-translated *EcR* was incubated with 1 ng ³²P-Hsp27-*EcRE* probe either alone or with *USP* or h*RXR-α*. MurA (Mur) or 20E were added as indicated and the resulting complexes



not shown) will respond to MurA when cotransfected with any of three *RXR*s or *USP*.

METHODS. *a*, Electrophoretic mobility shift assays with Hsp-27 *EcRE* probe were performed as described⁵ using proteins synthesized by coupled *in vitro* transcription/translation (Promega; TNT kit). In *b* and *c*, transfection and CAT assays were done as described⁵. After transfection, cells were treated with ethanol carrier (no hormone), 20 μM 20E (Sigma) or 1 μM MurA for 24–28 h before they were collected and analysed for CAT activity. CMX-*RXRα*, *β* and *γ* were constructed by inserting an *EcoRI* fragment containing the coding sequence of h*RXRα*, m*RXRβ*, or m*RXRγ* into CMXPL1 (K. Umehono, J.-D.C. and R.M.E., unpublished). Other plasmids have been described⁵.

► resolved by electrophoretic mobility shift assay. Affinity-purified polyclonal USP or crude RXR- α or EcR antiserum was included as indicated. The EcR/RXR- α DNA-binding activity is not seen unless the MurA is present (compare lanes 7 and 9; complex is indicated by an arrowhead) and 20E only affects the EcR/hRXR- α binding weakly (lane 8). The DNA-binding activity of EcR/USP complex, which is not strictly ligand-dependent, can also be enhanced by MurA and 20E (lanes 16 and 17). Similarly, both MurA and 20E stimulate EcRE binding activity of the native EcR/USP complex derived from Schneider cell line 2 nuclear extracts (data not shown). The identity of the components of the heterodimer are confirmed by employing antisera to EcR, RXR- α or USP to supershift the complex (lanes 10, 11 and 18, 19). Muristerone A had no effect on binding of EcR/USP to the non-functional 11N-EcRE mutant^{5,21} (data not shown). The induction of binding activity by MurA is strictly EcR-dependent as RXR- α and USP do not respond by themselves (lanes 6 and 14).

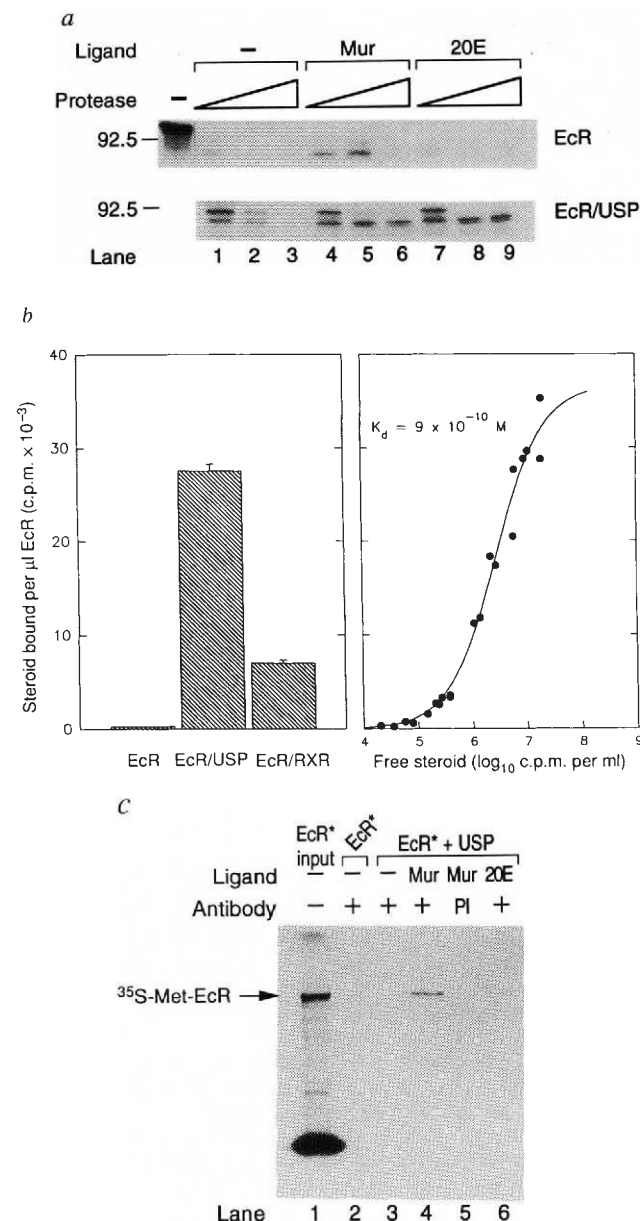
METHODS. Gel mobility shift assay was performed as described in Fig. 1 and ref. 5, with the following modification. Indicated proteins were preincubated with 1 μ M MurA or 20 μ M 20E for 30 min and then mixed with ³²P-Hsp27-EcRE probe and incubated for an additional 30 min at room temperature. Where indicated, antiserum was added for the final 15 min of the incubation. The affinity-purified polyclonal USP antibody has been described⁵. The EcR antibody (5377) was raised against a glutathione-S-transferase (GST) fusion protein containing amino acids 1–223 of EcR B1 (ref. 19) using previously described methods⁵.

FIG. 3 Ligand binding and EcR-USP association are interdependent. **a**, 20E and MurA induced protease resistance of EcR is enhanced by USP. In the top panel, ³⁵S-methionine-labelled EcR (EcR*) was subjected to proteolysis in the absence or presence of 1 μ M MurA or 20 μ M 20E. The bottom panel shows the results of a parallel experiment in which 60 ng purified GST-USP fusion protein⁵ was added before chymotrypsin cleavage. In the presence of USP, EcR* becomes more resistant to protease treatment in response to both 20E and MurA. Note that USP by itself can protect EcR from cleavage at low protease concentrations (compare lane 2 in each panel), suggesting that EcR and USP interact in solution (also, see **c**). Labelled EcR input precedes lane 1 (top). **b**, left panel: *in vitro*-translated EcR binding to ¹²⁵I-iodoponasterone is stimulated by USP or RXR- α . Binding reactions contained $\sim 3.7 \times 10^{-9}$ M ¹²⁵I-iodoponasterone (specific activity 1,697 Ci mmol⁻¹). Each bar represents mean ($\pm$ s.e.) specific binding per μ l of EcR of six (EcR) or eight (EcR/USP, EcR/RXR- α) determinations. Specific binding by EcR alone (246 ± 15 c.p.m. per μ l) was ~ 15 -fold higher than nonspecific background. Binding by EcR/USP and EcR/RXR- α is $27,552 \pm 752$ and $6,941 \pm 366$, respectively. Neither USP alone nor RXR- α alone gave detectable specific binding. Right, EcR/USP is similar to native ecdysone receptor in its affinity for ¹²⁵I-iodoponasterone. Each point represents the specific binding of 0.5 μ l EcR and 3.5 μ l USP at various concentrations of ¹²⁵I-iodoponasterone. Nonspecific binding (1.9% to total c.p.m.) has been subtracted. The best-fit curve corresponds to $K_d = 2.77 \times 10^{-6}$ c.p.m. ml⁻¹ (s.e., 0.34×10^6 c.p.m. ml⁻¹) and maximal binding is 3.67×10^4 c.p.m. per μ l EcR (s.e., 410 c.p.m.). **c**, the USP/EcR solution heterodimer is stabilized in solution by ecdysones. ³⁵S-methionine-labelled EcR (EcR*) was incubated with USP in the presence or absence of 1 μ M MurA or 20 μ M 20E and immunoprecipitation was performed using affinity-purified USP antibody. The resulting complexes were collected by protein A-agarose beads (Oncogene Science), washed, electrophoresed through a 10% SDS-polyacrylamide gel and visualized by autoradiography. MurA strongly increased the efficiency of EcR* precipitation (lane 4); 20E stimulated only weakly (lane 6). Preimmune serum did not precipitate EcR/USP complex (lane 5). EcR* precipitation is dependent on USP, regardless of whether MurA is present (lane 2, and data not shown).

METHODS. **a**, Protease cleavage analysis was performed as described¹¹. ³⁵S-methionine-labelled EcR was synthesized in a wheat-germ extract (Promega) and treated with increasing amounts of chymotrypsin (40, 50 and 75 μ g ml⁻¹) in the absence or presence of ligands and USP as indicated. The resulting cleavage products were separated by electrophoresis through a 10% SDS-polyacrylamide gel. Only the principal protected bands are shown. **b**, Steroid binding was measured as described¹². Total lysate (4 μ l) generated from a TNT *in vitro* translation kit (Promega) was used in one reaction tube. After a 1 h incubation with ¹²⁵I-iodoponasterone at room temperature, aliquots were removed for determination of total and protein-bound c.p.m. Reaction tubes for

localized on ecdysone-responsive loci of polytene chromosomes. Moreover, we show that natural ecdysones selectively promote physical association between EcR and USP, and conversely, that high-affinity hormone binding requires both EcR and USP. Replacement of USP with retinoid X receptor produces heterodimers with distinct pharmacological and functional properties. These results redefine the ecdysone receptor as a dynamic complex whose activity may be altered by combinatorial interactions among subunits and ligand.

We and others have previously shown that *Drosophila* USP dimerized with EcR and can substitute for its vertebrate counter-



measurements of nonspecific binding contained, in addition, unlabelled muristerone A (7×10^{-5} M final concentration) and the measured counts were subtracted to obtain specific binding. All samples were counted on a Packard γ -counter at $\sim 80\%$ efficiency. **c**, *In vitro*-translated EcR (³⁵S-methionine labelled) and USP were incubated for 1 h in DNA-binding buffer⁵ in the presence or absence of ligand. Immune/preimmune serum was added and incubated for a further 1 h. Protein A-agarose beads were added in 10 mM HEPES, pH 7.5, 150 mM NaCl, 1% NP-40 and deoxycholate, 1 mM dithiothreitol and 1 mM PMSF (RIPA buffer) and incubated at 4 $^{\circ}$ C for 1 h. The beads were collected by centrifugation, washed 5 times with RIPA buffer and the precipitated proteins analysed on a 10% SDS-polyacrylamide gel.

part, the retinoid X receptor (RXR)⁶, to form high-affinity DNA-binding complexes with several vertebrate nuclear receptors^{5,7,8}. Although EcR/USP binds specifically to an ecdysone-response element (Hsp27-EcRE⁹), we unexpectedly found that no binding could be detected with EcR or the retinoid X α -receptor (RXR- α) (Fig. 1a). Functionally, USP can interact with the EcR and a transcriptionally more active derivative, GEcR⁵ (Fig. 1b and legend) to confer 20-hydroxy-ecdysone (20E) responsiveness to a heterologous cell line CV1. But co-transfection with RXR- α fails to promote ecdysone responsiveness (Fig. 1b). Although EcR alone cannot transactivate in CV1 cells in response to 20E, EcR expressed in mammalian 293 cells can respond to muristerone A (MurA), a potent ecdysone agonist¹⁰. This paradox prompted us to compare the activities of these two hormones in activating GEcR. As with 20E, GEcR alone expressed in CV1 cells does not respond to MurA (Fig. 1c). In contrast, MurA can activate GEcR by coexpressing either USP or RXR- α , - β or - γ (Fig. 1c). Thus GEcR/USP and GEcR/RXR α s are commonly activated by MurA, but are differentially responsive to 20E. These data suggest that ligand responsiveness within the functional EcR complex is modulated by the EcR partner.

Mobility shift assays were used to explore the effects of MurA and 20E on DNA binding. EcR and RXR- α do not bind to the

EcRE probe either individually or when mixed (Fig. 2, lanes 1, 4 and 7). But EcR/RXR- α DNA binding is dramatically enhanced by MurA (lane 9), as was also shown previously⁸. In contrast, 20E only stimulates binding weakly (lane 8). These data indicate that EcR/USP (but not EcR-RXR) binds the EcRE even in the absence of ligand (lane 15) and that these complexes are both enhanced by MurA (lanes 9, 17) but selectively by 20E (compare lanes 8 and 16), which is consistent with the transfection results (Fig. 1b and c).

Together these data suggested that the ligand may be an allosteric effector of heterodimer formation and led us to the reciprocal speculation that heterodimer formation may be required for hormone binding. A protease protection assay was used to investigate this question because nuclear hormone receptors typically show ligand-induced resistance to protease cleavage¹¹. As shown in Fig. 3a (upper panel), ³⁵S-labelled EcR alone is cleaved by increasing amounts of chymotrypsin (lanes 1–3). MurA treatment leads to a set of protected fragments that are increasingly resistant to protease cleavage (upper panel, lanes 4 and 5). In contrast, 20E is ineffective in promoting protease resistance (lanes 7–9), suggesting that in the absence of a partner EcR can bind to MurA but not 20E. In the presence of USP (Fig. 3a, lower panel), MurA and 20E induce protease resistance even at high chymotrypsin concentrations (compare lanes 6 and 9 in

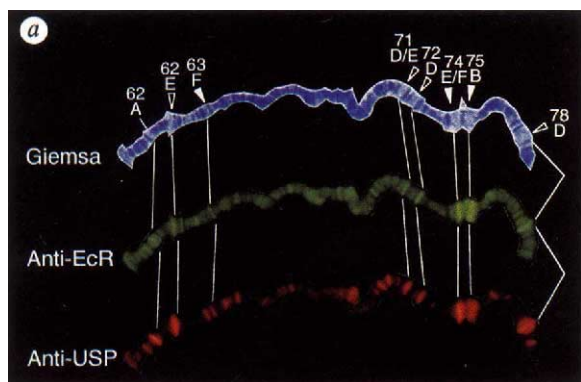
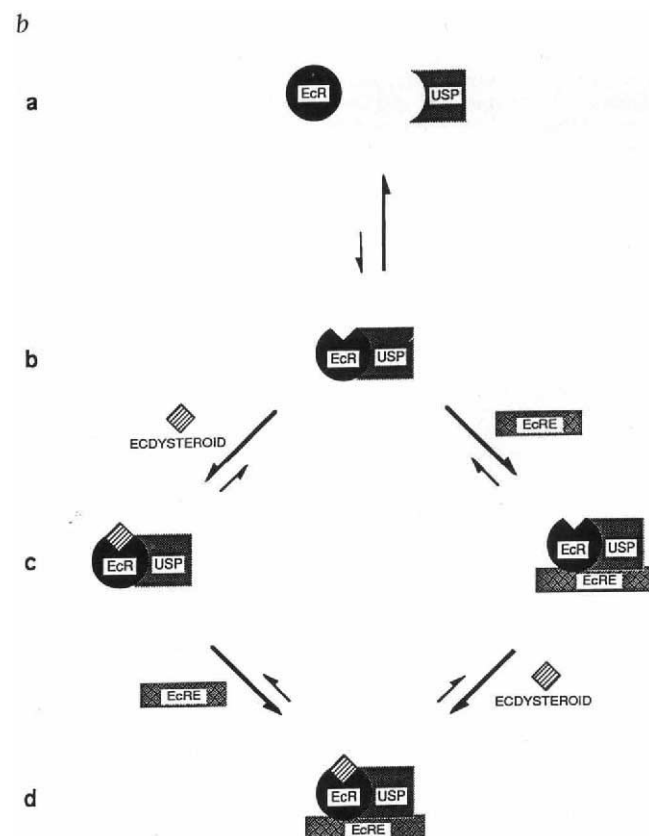


FIG. 4 a, EcR and USP are colocalized on ecdysone-responsive loci of polytene chromosomes. Salivary gland chromosome 3L prepared from late third instar larva was stained with Giemsa (top) and subjected to simultaneous immunological staining with antibodies recognizing EcR (middle) or USP (bottom). Chromosome puffs can be seen on the early ecdysone responsive loci 74E/F and 75B (filled arrow) and early-late responsive loci 62E, 71D/E, 72D and 78D (open arrow). Puff on another early responsive locus 63F is less obvious. This chromosome is roughly at puffing state 7 according to the Ashburner's classification scheme². The binding sites of EcR and USP on chromosome 3L were determined by staining with EcR antibody (fluorescein-labelled secondary; green) and USP antibody (rhodamine labelled secondary; red). Signals can be detected on the chromosome by both EcR and USP antibody. Specifically, both EcR and USP binding can be colocalized to the ecdysone-responsive loci already mentioned and indicated by lines. **b**, Schematic model illustrating the interdependence of ligand binding and subunit association in the native EcR complex. In (a), the subunits are incapable of high-affinity ligand- or DNA-binding but are in equilibrium with the unstable EcR/USP heterodimer (b). Although the EcR/USP complex is not stable, this interaction presumably induces a conformation change in EcR (as indicated by a notch) and the complex becomes capable of high-affinity binding to ecdysones and to EcREs (c). Binding of ligand and/or EcREs to EcR/USP favours the formation of stable EcR/USP complexes to create a functional complex (d) that transduces the ecdysone signal.

METHODS. **a**, Procedures for immunostaining of the polytene chromosomes²² were followed, with minor modifications. Briefly, salivary glands from late third instar were dissected and fixed and squashed on a microscope slide. The slide was transferred to dry ice for 15 min, the coverslip removed, and then washed with PBS followed by PBS with



0.1% Triton X-100. The slides were blocked with PBS, 0.1% Triton X-100, 5% normal goat serum and then mouse monoclonal anti-EcR (DDA2.7)¹⁹ and rabbit polyclonal anti-USP antibodies⁵ were added at 1:4 and 1:80 dilution respectively. These primary antibodies were incubated at room temperature for 2 h, washed in PBS with 0.1% Triton X-100, and visualized with secondary antibodies as follows: rhodamine-conjugated goat anti-rabbit (Cappel) and biotinylated anti-mouse (Vector)/fluorescein-conjugated streptavidin (Vector). Chromosomes were stained with Giemsa (Gurr) to visualize the banding pattern.

two panels). Whereas USP is an essential partner for 20E binding, dose-response studies indicate that MurA affinity is also enhanced by USP (data not shown). These results were confirmed directly by ligand-binding experiments with the ecdysone analogue 26-¹²⁵I-iodoponasterone A (¹²⁵I-iodoponasterone)¹². EcR alone bound ligand weakly (Fig. 3b, left panel), but ¹²⁵I-iodoponasterone binding was dramatically stimulated by addition of USP. The estimated K_d for EcR/USP complex binding to ¹²⁵I-iodoponasterone is 9×10^{-10} M (Fig. 3b, right panel), close to the K_d estimated for the endogenous ecdysone receptor from *Drosophila* nuclear extract¹². A similar result on direct ligand binding has also been reported¹³. RXR- α can also enhance EcR ligand binding, although less effectively. No ligand binding was detected with either USP or RXR- α alone (data not shown). These results show that high-affinity hormone binding is an intrinsic property of the EcR/USP complex but not of the individual components. This is in marked contrast to the mammalian retinoic acid receptor, thyroid hormone receptor and vitamin D receptor, whose high-affinity ligand binding is not apparently altered by RXR (refs 14, 15; and J. W. Pike, personal communication).

The unusual property of EcR/USP in binding ecdysone and its dependence on ligand to achieve maximal DNA binding suggest that ligand may function to reduce the dissociation of a transient EcR-USP complex. We therefore used immunoprecipitation assays to probe for solution interactions between EcR and USP. As shown in Fig. 3c, when labelled EcR is incubated with USP, a very small amount of EcR can be co-precipitated by anti-USP antibody (lane 3). Inclusion of MurA greatly enhances co-precipitation, but 20E does so only weakly (lanes 4 and 6, respectively). These results show that EcR and USP form an unstable complex in solution which is stabilized by ligand. Stabilization of EcR/USP in solution may contribute to the higher DNA-binding activity shown by the EcR/USP-ligand complex (Fig. 2).

The biochemical data would suggest that native ecdysone-responsive complex *in vivo* should contain both EcR and USP. To confirm this, we used immunohistochemistry to localize the position of EcR and USP on polytene chromosomes. Indeed, as illustrated with chromosome 3L, USP and EcR colocalize to the early ecdysone inducible puffs E74 (refs 16, 17) and E75 (ref. 18), as well as other known major ecdysone-responsive loci (see Fig. 4a and legend for details). These data demonstrate that EcR/USP complexes are present at ecdysone-responsive loci *in vivo* and strongly support that EcR/USP heterodimer is the signal-transducing complex for ecdysone.

Our observations indicate that the functional ecdysone receptor is the product of both the EcR and USP loci. These products associate weakly in solution (Fig. 4b) to form a functional complex capable of high-affinity ligand binding. Ligand, in turn, shifts the equilibrium towards a stable complex. This novel arrangement redefines the ecdysone receptor as a ternary complex displaying interdependent interactions between ligand and the EcR/USP subunits. Our model suggests that the ecdysone receptor may represent a dynamic complex whose properties are defined by the precise subunit composition in conjunction with specific ligands. As EcR/USP and EcR/RXR exhibit differential responses, we speculate that partners other than USP may exist that could confer differential responsiveness to the many known naturally occurring ecdysones. □

Received 23 August; accepted 5 November 1993.

1. Clever, U. & Karlson, P. *Exp. Cell Res.* **20**, 623–626 (1960).
2. Ashburner, M. in *Developmental Studies on Giant Chromosomes* (ed. Beermann, W.) 101–147 (Springer, New York, 1972).
3. Ashburner, M., Chihara, C., Meltzer, P. & Richards, G. *Cold Spring Harbor Symp. quant. Biol.* **38**, 655–662 (1974).
4. Koelle, M. R. et al. *Cell* **67**, 59–77 (1991).
5. Yao, T.-P., Segraves, W. A., Oro, A. E., McKeown, M. & Evans, R. M. *Cell* **71**, 63–72 (1992).
6. Mangelsdorf, D. J. et al. *Genes Dev.* **6**, 329–344 (1992).
7. Christianson, A. M. K. et al. *Proc. natn. Acad. Sci. U.S.A.* **89**, 11503–11507 (1992).
8. Thomas, H. E., Stunnenberg, H. G. & Stewart, A. F. *Nature* **362**, 471–475 (1993).

9. Riddihough, G. & Pelham, H. R. B. *EMBO J.* **6**, 3729–3734 (1987).
10. Christopherson, K. S., Mark, M. R., Bajaj, V. & Godowski, P. J. *Proc. natn. Acad. Sci. U.S.A.* **89**, 6314–6318 (1992).
11. Allan, G. F. et al. *J. biol. Chem.* **267**, 19513–19520 (1992).
12. Cherbas, P., Cherbas, L., Lee, S.-S. & Nakanishi, K. *Proc. natn. Acad. Sci. U.S.A.* **85**, 2096–2100 (1988).
13. Koelle, M. R., Arbeitman, M. & Hogness, D. S. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
14. Yang, N., Schüle, R., Mangelsdorf, D. M. & Evans, R. M. *Proc. natn. Acad. Sci. U.S.A.* **88**, 3559–3563 (1991).
15. Forman, B. M., Casanova, J., Raaka, B. M., Ghysdael, J. & Samuels, H. H. *Molec. Endocr.* **6**, 429–442 (1992).
16. Burtis, K. C., Thummel, C. S., Jones, C. W., Karim, F. D. & Hogness, D. S. *Cell* **61**, 85–99 (1990).
17. Thummel, C. S., Burtis, K. C. & Hogness, D. S. *Cell* **61**, 101–111 (1990).
18. Segraves, W. A. & Hogness, D. S. *Genes Dev.* **4**, 204–219 (1990).
19. Talbot, W. S., Swyryd, E. A. & Hogness, D. S. *Cell* **73**, 1323–1337 (1993).
20. Hollenberg, S. M. & Evans, R. M. *Cell* **55**, 899–906 (1988).
21. Cherbas, L., Lee, K. & Cherbas, P. *Genes Dev.* **5**, 120–131 (1991).
22. Zink, B., Engström, Y., Gehring, W. J. & Paro, R. *EMBO J.* **10**, 153–172 (1991).

ACKNOWLEDGEMENTS. We thank D. Hogness, M. Koelle, M. Arbeitman and J. W. Pike for communicating their results before publication; D. Hogness and W. Talbot for anti-EcR monoclonal antibody; J. Thomas for advice on polytene chromosome staining; J. Dyck for RXR α antibody; and T. Perlmann, A. Oro, B. Blumberg and members of R.M.E.'s laboratory for critically reading the manuscript and for suggestions. This work was supported by the Howard Hughes Medical Institute (R.M.E., T.-P.Y. and J.D.C.), Jane Coffin Childs Memorial Fund for Cancer Research or the Tobacco Related Disease Research Program (B.M.F.), Dupont (P.C., L.C. and Z.J.), and grants from the NIH (R.M.E., M.M., P.C., L.C. and Z.J. M.M.).

Chimaeric nicotinic-serotonergic receptor combines distinct ligand binding and channel specificities

Jean-Luc Eiselé*, Sonia Bertrand†, Jean-Luc Galzi*, Anne Devillers-Thiéry*, Jean-Pierre Changeux* & Daniel Bertrand†

* Neurobiologie Moléculaire, Unité de Recherche Associée au Centre National de la Recherche Scientifique D1284, Institut Pasteur, 25 rue du Dr Roux, 75724 Paris Cedex 15, France

† Département de Physiologie, Centre Médical Universitaire (Faculté de Médecine), 1211 Geneva 4, Switzerland

THE neuronal nicotinic $\alpha 7$ (nAChR) and 5-hydroxytryptamine (5HT₃) receptors^{1–3} are ligand-gated ion channels with a homologous topological organization and have activation and desensitization reactions in common. Yet these homo-oligomeric receptors differ in the pharmacology of their binding sites for agonists and competitive antagonists^{3,4}, and in their sensitivity to Ca²⁺ ions. The $\alpha 7$ channel is highly permeable to Ca²⁺ ions^{5,6} and external Ca²⁺ ions potentiate, in an allosteric manner, the permeability response to acetylcholine, as shown for other neuronal nAChRs^{7,8}. The 5HT₃ channel, in contrast, is not permeable to Ca²⁺ ions, but blocked by them^{3,9}. To assign these properties to delimited domains of the primary structure, we constructed several recombinant chimaeric $\alpha 7$ -5HT₃ receptors. We report here that one of the constructs expresses a functional receptor that contains the serotonergic channel still blocked by Ca²⁺ ions, but is activated by nicotinic ligands and potentiated by external Ca²⁺ ions.

The neurotransmitter receptors from the superfamily of ligand-gated ion channels are composed of several transmembrane subunits that share a common organization of their hydrophilic and hydrophobic domains. In the nicotinic, glycine and GABA_A receptors, affinity labelling and site-directed mutagenesis reveal that the large extracellular amino-terminal segment forms at least three loops that contribute to the neurotransmitter binding domain^{10–15}. Four 25–30-amino-acid stretches (M1 to M4) probably span the membrane^{16–18} and M2 lines the ion

channel¹⁹⁻²¹. To correlate the functional properties of the neurotransmitter binding site and ion channel to defined structural domains, we constructed recombinant chimaeric subunits containing the N-terminal part of the nicotinic $\alpha 7$ subunit and the complementary C-terminal domain of the serotonergic 5HT₃ subunit. Five different complementary DNA constructs were made to join fragments of the $\alpha 7$ subunit and of the 5HT₃ subunit at conserved residues tryptophan at position 173 (W173),

Y194, V201, L208 and P217 (Fig. 1, top). Both $\alpha 7$ and 5HT₃ wild type subunits form functional homo-oligomeric receptors on heterologous expression in *Xenopus* oocytes^{1,3}. With the exception of $\alpha 7$ -W173-5HT₃, which lacks the third binding site loop, all chimaeric cDNA constructs yielded nicotine-sensitive ¹²⁵I-labelled α -bungarotoxin binding sites on expression in HEK 293 cells. This indicates that the large hydrophilic N-terminal domain of $\alpha 7$ (residues 1 to at least 194) folds correctly in the

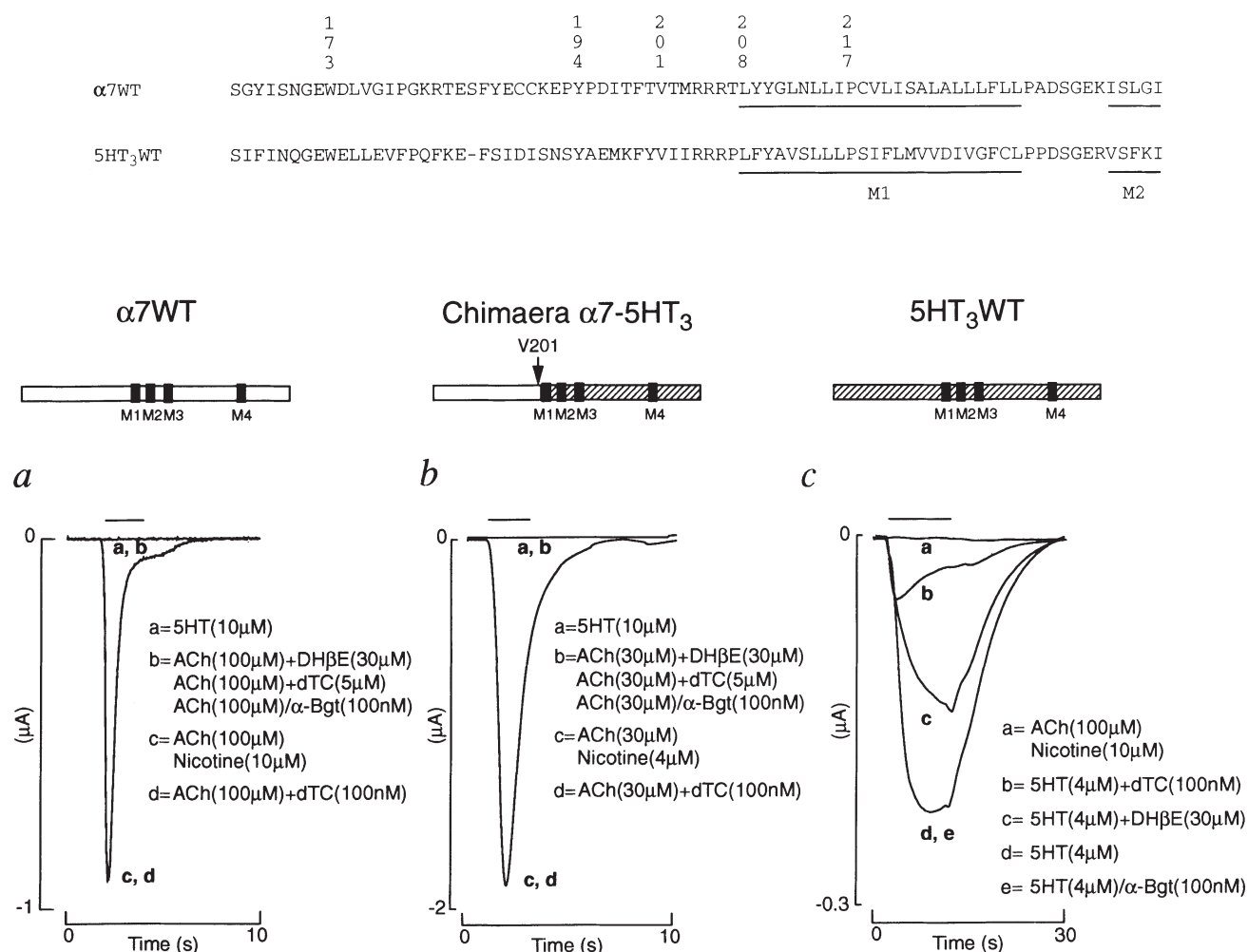


FIG. 1 Construction of chimaeric receptors and pharmacological characterization. Top, amino-acid sequence (single-letter code) comparison of $\alpha 7$ WT and 5HT₃WT. Conserved positions numbered according to $\alpha 7$ sequence, W173, Y194, V201, L208, P217, were used as junction points for the construction of the recombinant chimaeric receptors. a–c, Time course evoked currents in *Xenopus* oocyte from $\alpha 7$ WT (a), $\alpha 7$ -V201-5HT₃ chimaera (b) and 5HT₃WT (c). Agonists: acetylcholine (ACh), nicotine, 5-hydroxytryptamine (5HT); and antagonists: dihydro- β -erythroidine (DH β E), tubocurarine (dTC), α -bungarotoxin (α -Bgt) were applied for 2 s ($\alpha 7$ WT and chimaera) or 10 s (5HT₃WT) (horizontal bars). Antagonism by dTC or DH β E was tested by simultaneous application with the agonist. α -Bgt incubations were as follows: cells were first challenged with the appropriate agonist and then incubated in 60 μ l of OR2 medium (see below) containing 100 nM α -Bgt and 20 μ g ml⁻¹ BSA. After a 30 min incubation the oocytes were placed in the recording chamber and challenged again with the agonist. Cells were held throughout the experiment at -100 mV.

METHODS. $\alpha 7$ cDNA was provided by M. Ballivet. 5HT₃ cDNA was cloned by polymerase chain reaction (PCR). The genes were then transferred into pCDNA as HindIII–XbaI fragments, with a consensus translation initiation site. Chimaeric receptors were constructed by two sequential

PCR³³, cloned as EcoRI–XbaI or HindIII–XbaI ($\alpha 7$ -W173-5HT₃) fragments into pCDNA- $\alpha 7$, sequenced and injected into oocytes as described below. We found that chimaera constructed into a pMT3 vector³⁴ leads larger agonist-evoked currents and was therefore used in pharmacological experiments. Electrophysiology: oocytes were prepared, injected and recorded as described³⁵. Recordings were typically made two or three days after injection. For voltage-clamp measurements, cells were incubated in OR2 medium (NaCl, 82.5 mM; KCl, 2.5 mM; CaCl₂, 2.5 mM; MgCl₂, 1 mM; atropine, 0.5 μ M; HEPES, 5 mM, pH 7.4). Dose–response curves were obtained by plotting the peak of the ACh-evoked currents, measured for 2-s applications, as a function of the logarithm of the agonist concentration. Cells were held at -100 mV. Mean currents obtained in 5 cells from several donors expressing each mutant were normalized to the values determined at saturating agonist concentration. The empirical Hill equation, adjusted by weighted least-square algorithm, yielded the apparent 50% effective concentrations (EC₅₀) and Hill coefficients given in the text. IC₅₀ for competitive antagonists were determined in the presence of agonist at its EC₅₀ concentration. To prevent activation of chloride currents by accumulation of intracellular calcium ions, the calcium chelator BAPTA was injected as described previously³⁰.

chimaeric subunit to form a pharmacological binding site for cholinergic ligands, provided that all of the amino acids identified by affinity labelling and mutagenesis are conserved. After injection in *Xenopus* oocytes, chimaeras $\alpha 7$ -W173-5HT₃, $\alpha 7$ -L208-5HT₃ and $\alpha 7$ -P217-5HT₃ did not exhibit any detectable ionic currents (less than 1 nA) to 1 mM acetylcholine (ACh) application, indicating that the requirements for functional coupling appear more stringent than for ligand binding^{22–24}. In contrast, two chimaeras expressed functional receptors of small (5–10 nA for $\alpha 7$ -Y194-5HT₃) and large (30–300 nA for $\alpha 7$ -V201-5HT₃) amplitude of ACh-evoked currents, thus pointing to a functional coupling between the cholinergic ligand-binding domain of the $\alpha 7$ subunit and the channel domain of the 5HT₃ subunit in these chimaeras.

The pharmacological properties of $\alpha 7$ and 5HT₃ wild-type receptors (WT) and of the $\alpha 7$ -V201-5HT₃ chimaera are illustrated in Fig. 1. On the $\alpha 7$ WT receptor, acetylcholine (50% effective concentration, EC₅₀ = 115 μ M) and nicotine (EC₅₀ = 10.5 μ M) are agonists, whereas dihydro- β -erythroidine (DH β E; 50% inhibitory concentration, IC₅₀ = 1.6 μ M), curare (dTC; IC₅₀ = 0.5 μ M) and α -bungarotoxin (α -Bgt; IC₅₀ = 0.7 nM) are

antagonists^{1,4}. Serotonin (5HT) neither behaves as an agonist nor an antagonist on the $\alpha 7$ WT receptor.

By contrast, 5HT (EC₅₀ = 3.6 μ M) is an agonist and curare is a high-affinity antagonist (IC₅₀ = 2–5 nM) of both the endogenous and heterologously expressed 5HT₃WT receptor^{3,25–28}. Moreover, we found that the competitive antagonist DH β E inhibits the 5HT₃WT receptor with a low affinity (IC₅₀ = 48 μ M). ACh, nicotine and α -Bgt (100 nM) have no effect by themselves or on 5HT-evoked responses (refs 3, 28, and this work).

On the $\alpha 7$ -V201-5HT₃ chimaera, ACh and nicotine, but not 5HT, are agonists with EC₅₀ values of 25 μ M and 4.2 μ M, respectively. Furthermore, α -Bgt, DH β E (IC₅₀ = 1.5 μ M) and curare at high concentration (5 μ M) are antagonists of the chimaera, indicating that the determined pharmacological profile is closely related to that of a nicotinic receptor and suggesting a negligible, if any, contribution of the C-terminal sequence to the neurotransmitter binding area.

A significant difference between the nicotinic and the 5HT₃ receptors concerns their sensitivity to the presence of external calcium. As shown in Fig. 2, the amplitude of agonist-evoked $\alpha 7$ WT currents increases with increasing external calcium con-

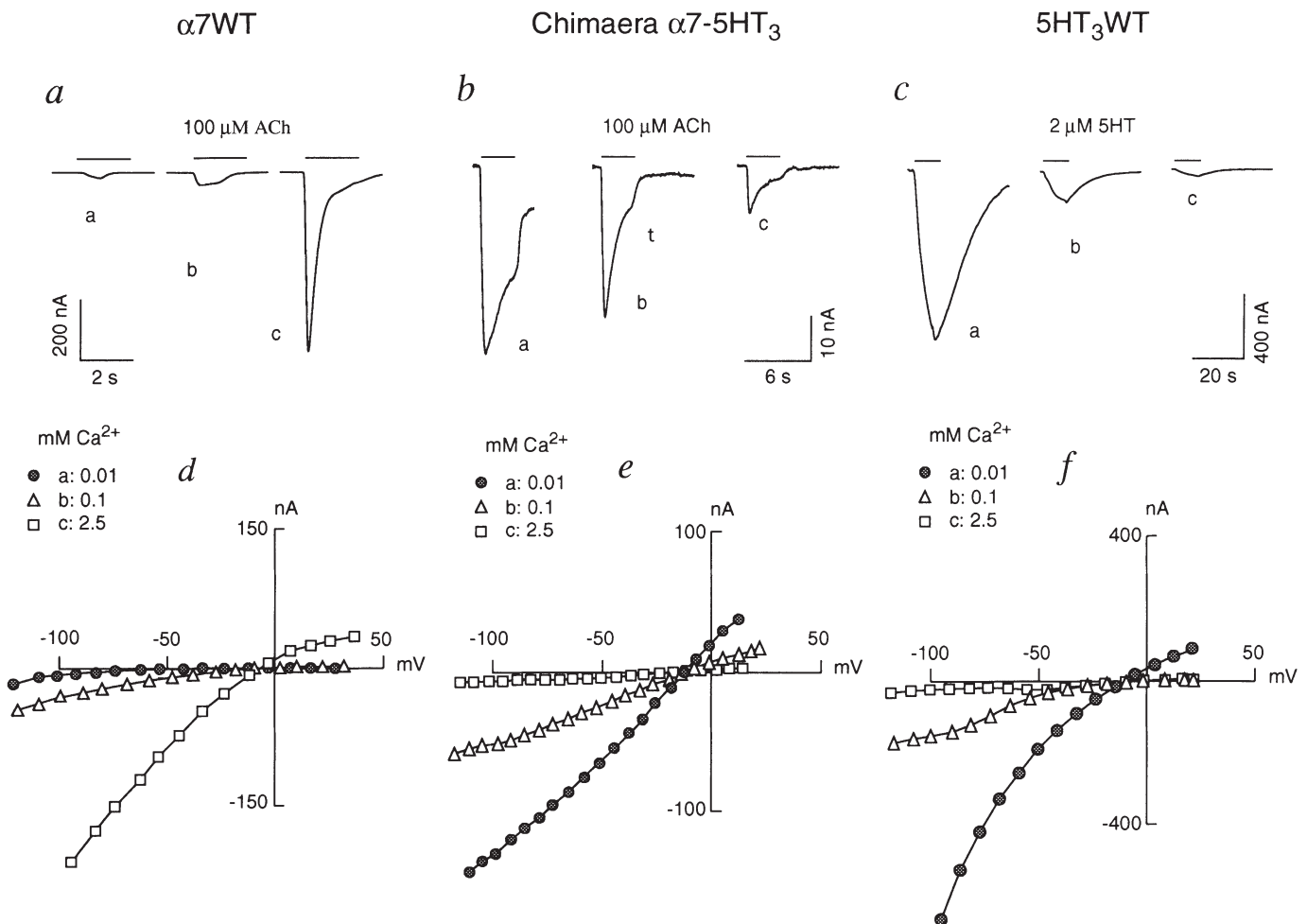


FIG. 2 Influence of the extracellular calcium concentration on the time course and amplitudes of responses and on current-voltage relationships. *a–c*, Responses to a fixed agonist concentration, indicated on the top, were recorded in three external calcium concentrations (a, 0.01 mM; b, 0.1 mM; c, 2.5 mM). Cells were held at -100 mV. Note the difference in timescale between the three receptors. *d–f*, Current-voltage relationships were obtained by application of a voltage ramp

during a steady agonist concentration and generated by subtracting passive membrane currents from agonist-evoked currents. Fast ramps going from -140 to $+40$ mV synchronized on the peak of the ACh evoked current were used in *d*. To avoid partial calcium channel block, *e* and *f* were obtained with slower ramps going from $+40$ mV to -140 mV; *d–f* were recorded with the same agonist concentration as *a–c*.

centration. This effect is in part due to a potentiation of ACh responses (see below) and in part to the high calcium permeability of the ion channel^{5,6} that leads to an elevation of intracellular calcium and subsequent activation of calcium-dependent chloride channels in the oocyte^{8,29,30}. By contrast, both on endogenous and on heterologous 5HT₃ receptors^{3,31}, the amplitude of 5HT-evoked current on the 5HT₃WT receptor declines when external calcium concentration increases. This suggests the occurrence of a channel block by calcium. The inward rectification observed on current-voltage relationships (at 0.1 and 2.5 mM calcium) further indicates that this effect is voltage-dependent. ACh-evoked currents on the chimera are also reduced, in a voltage-dependent manner, when external calcium concentration is raised from 0.01 to 2.5 mM, thus indicating that, as in the 5HT₃WT receptor, the channel domain of the chimera is blocked by calcium.

In the absence of chloride contribution³⁰, agonist-evoked α 7WT currents are still potentiated by external calcium (Fig. 3a). This property, shared by several neuronal nicotinic receptors^{7,8}, has been attributed to calcium binding to an external site which affects the open probability of the ion channel. A small but significant increase of the EC₅₀ of ACh from 375 μ M at 0.01 mM external calcium to 115 μ M at 2.5 mM external calcium (comparable to nAChRs from cultured medial habenular neurons⁷) is observed (Fig. 3d). In contrast, modification of the external calcium concentration does not significantly affect the EC₅₀ for 5HT on the 5HT₃WT receptor (data not shown). Also, a rapid switch from low to high calcium concentration during a single pulse of 5HT application (Fig. 3c), a protocol expected to resolve a rapid potentiating effect⁷ from a slower channel block, only results in a reduction of the current due to channel block without any sign of potentiation. The same protocol of

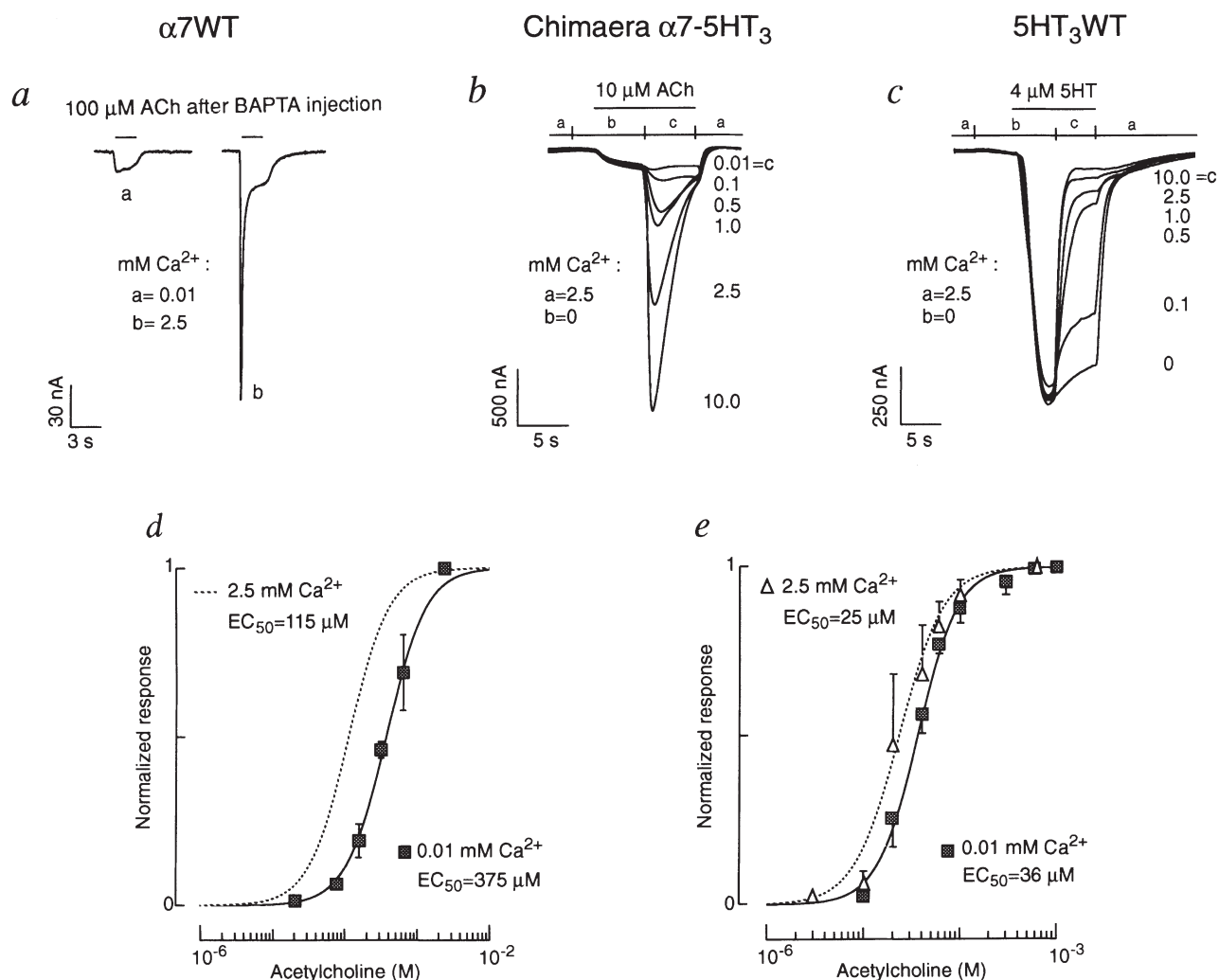


FIG. 3 Potentiation and channel block caused by extracellular calcium. **a**, Responses to 100 μ M ACh were recorded in the same cell, at a holding potential of -100 mV, in low and normal calcium concentration. To prevent any contamination by chloride currents, responses were recorded several minutes after intracellular injection of the calcium-chelating agent BAPTA (see Fig. 1 methods). **b**, Calcium both potentiates currents and blocks the ion channel of the chimaeric receptor. A dual application protocol was defined to reveal the effects of extracellular calcium. Control medium (**a**) was first replaced by a solution containing no calcium (**b**) and a fixed concentration of agonist was applied (10 μ M) indicated by the upper bar. Calcium was then abruptly increased in the extracellular medium (**c**). Traces recorded in the same cell for different calcium concentrations are superimposed. The cell was maintained

during the entire experiment at -100 mV. **c**, Extracellular calcium blocks the 5HT₃WT receptor channel. Application of the dual protocol experiment (as in **b**) on an oocyte displaying large 5HT-evoked current reveals only a channel block. **d**, Reduction of the extracellular calcium concentration provokes a right shift of the ACh dose-response curve in the α 7WT receptor. Dose-response curve (see Fig. 1 methods) were measured in control (dashed line) or low calcium concentration (continuous line and filled squares). **e**, Peak currents in cells expressing chimaeric receptors were measured as in **d** at two calcium concentrations. In **d** and **e**, lines were drawn using the empirical Hill equations with the indicated EC₅₀. Hill coefficients are 1.6 for the α 7WT and 1.8 for the chimera.

rapid switch from low to high external calcium concentration applied to the chimaera first results in a potentiation of the ionic current, followed by a decline of response amplitude due to channel block (Fig. 3b). Also, in support of the potentiating effect of calcium on the chimaera, a slight increase of EC_{50} for ACh from 36 μ M at 0.01 mM external calcium to 25 μ M at 2.5 mM external calcium is noticed (Fig. 3e). These data show that the large N-terminal domain of the nicotinic $\alpha 7$ receptor not only contains the neurotransmitter binding site, but also the allosteric site(s) for calcium.

Thus, functional properties that are expected to be determined by the extracellular domain on one hand, or by the channel domain on the other hand, are found to be associated with the corresponding protein segment in the chimaera. In contrast, the kinetics of current onset and desensitization, which are both rapid in $\alpha 7$ WT and slow in 5HT₃WT, are intermediate in the chimaera (Figs 1 and 2). This suggests that the rates of inter-conversion between the multiple conformational states of a given ligand-gated ion channel (for both activation and desensitization) involve the tertiary and quaternary structure of the whole receptor molecule rather than a restricted structural domain.

Distinct functional properties of two members among the superfamily of the ligand-gated ion channels can thus be combined into a new chimaeric $\alpha 7$ -5HT₃ receptor by the exchange of homologous sequence elements. This allows us to assign, without ambiguity, neurotransmitter binding and Ca^{2+} allosteric properties to the large N-terminal hydrophilic domain, and ionic channel properties to the other moiety. But the successful coupling of the neurotransmitter binding site from one receptor to the ion channel from the other supports the additional notion of at least two independent protein domains with folding autonomy³² and functional specificity, which behave as elementary units within the ligand-gated channel complex. Moreover, their ability to interact functionally, despite the low sequence homology of the receptors from which they originate, suggests that the different members of the ligand-gated ion channels superfamily share a highly conserved architectural framework designed to mediate signal transduction. □

Received 25 August; accepted 5 October 1993.

1. Couturier, S. et al. *Neuron* **5**, 845–856 (1990).
2. Schoepfer, R., Conroy, W. G., Whiting, P., Gore, M. & Lindstrom, J. *Neuron* **5**, 35–48 (1990).
3. Maricq, A. V., Peterson, A. S., Brake, A. J., Myers, R. M. & Julius, D. *Science* **254**, 432–437 (1991).
4. Bertrand, D., Bertrand, S. & Ballivet, M. *Neurosci. Lett.* **146**, 87–90 (1992).
5. Séguela, P., Wadiche, J., Dineley-Miller, K., Dani, J. A. & Patrick, J. W. *J. Neurosci.* **13**, 596–604 (1993).
6. Bertrand, D., Galzi, J. L., Devillers-Thiéry, A., Bertrand, S. & Changeux, J. P. *Proc. natn. Acad. Sci. U.S.A.* **90**, 6971–6975 (1993).
7. Mülle, C., Léna, C. & Changeux, J. P. *Neuron* **8**, 937–945 (1992).
8. Vernino, S., Amador, M., Luetje, C. W., Patrick, J. & Dani, J. A. *Neuron* **8**, 127–134 (1992).
9. Peters, J. A., Hales, T. G. & Lambert, J. J. *Eur. J. Pharm.* **151**, 491–495 (1988).
10. Dennis, M. et al. *Biochemistry* **27**, 2346–2357 (1988).
11. Galzi, J. L. et al. *J. biol. Chem.* **265**, 10430–10437 (1990).
12. Kuhse, J., Schmieden, V. & Betz, H. *Neuron* **5**, 867–873 (1990).
13. Vandenberg, R. J., French, C. R., Barry, P. H., Shine, J. & Schofield, P. R. *Proc. natn. Acad. Sci. U.S.A.* **89**, 1765–1769 (1992).
14. Sigel, E., Baur, R., Kellenberger, S. & Malherbe, P. *EMBO J.* **11**, 2017–2023 (1992).
15. Karlin, A. *Curr. Opin. Neurobiol.* **3**, 299–309 (1993).
16. Claudio, T., Ballivet, M., Patrick, J. & Heinemann, S. *Proc. natn. Acad. Sci. U.S.A.* **80**, 1111–1115 (1983).
17. Devillers-Thiéry, A., Giraudat, J., Bentabollet, M. & Changeux, J. P. *Proc. natn. Acad. Sci. U.S.A.* **80**, 2067–2071 (1983).
18. Noda, M. et al. *Nature* **301**, 251–255 (1983).
19. Giraudat, J., Dennis, M., Heidmann, T., Chang, J. Y. & Changeux, J. P. *Proc. natn. Acad. Sci. U.S.A.* **83**, 2719–2723 (1986).
20. Hucho, F., Oberthür, W. & Lottspeich, F. *FEBS Lett.* **205**, 137–142 (1986).
21. Imoto, K. et al. *Nature* **324**, 670–674 (1986).
22. Mayne, K. M., Yoshii, K., Yu, L., Lester, H. A. & Davidson, N. *Molec. Brain Res.* **2**, 191–197 (1987).
23. Yu, X. M. & Hall, Z. W. *Nature* **352**, 64–67 (1991).
24. Sumikawa, K. *Molec. Brain Res.* **13**, 349–353 (1992).
25. Bradley, P. B. et al. *Neuropharmacology* **25**, 563–576 (1986).
26. Higashi, H. & Nishi, S. *J. Physiol., Lond.* **323**, 543–567 (1982).
27. Yakel, J. L., Lagrutta, A., Adelman, J. P. & North, R. A. *Proc. natn. Acad. Sci. U.S.A.* **90**, 5030–5033 (1993).
28. Yang, J., Mathie, A. & Hille, B. *J. Physiol., Lond.* **448**, 237–256 (1992).
29. Miledi, R. & Parker, I. J. *Physiol.* **357**, 173–183 (1984).
30. Galzi, J. L. et al. *Nature* **359**, 500–505 (1992).
31. Yang, J. J. *Gen. Physiol.* **96**, 1177–1198 (1990).
32. Popot, J. L. & Engelman, D. *Biochemistry* **29**, 4031–4037 (1990).
33. Higuchi, R., Krummel, B. & Saiki, R. K. *Nucleic Acids Res.* **16**, 7351–7367 (1989).
34. Swick, A. G., Janicot, M., Cheneval-Kastelic, T., McLenithan, J. C. & Lane, M. D. *Proc. natn. Acad. Sci. U.S.A.* **89**, 1812–1816 (1992).
35. Bertrand, D., Cooper, E., Valera, S., Rungger, D. & Ballivet, M. *Electrophysiology of Neuronal Nicotinic Acetylcholine Receptors Expressed in Xenopus Oocytes Following Nuclear Injection of Genes or cDNA 1-174-193* (Academic, San Diego, 1991).

ACKNOWLEDGEMENTS. This work was supported by grants from the Association Française contre les Myopathies, the Collège de France, the Centre National de la Recherche Scientifique, the Ministère de la Recherche, the Direction des Recherches Etudes et Techniques, the Commission of the European Communities, the International Human Frontier Science Program, and the Swiss National Science Foundation, OFES and Sandoz Foundation (to D.B.). J.L.G. is a recipient of a short-term EMBO fellowship.

The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition

Louise Fairall, John W. R. Schwabe,
Lynda Chapman, John T. Finch
& Daniela Rhodes

MRC Laboratory of Molecular Biology, Hills Road,
Cambridge CB2 2QH, UK

THE Cys₂-His₂ zinc-finger is the most widely occurring DNA-binding motif^{1–3}. The first structure of a zinc-finger/DNA complex revealed a fairly simple mechanism for DNA recognition⁴ suggesting that the zinc-finger might represent a candidate template for designing proteins to recognize DNA⁵. Residues at three key positions in an α -helical 'reading head' play a dominant role in base-recognition and have been targets for mutagenesis experiments aimed at deriving a recognition code^{6–8}. Here we report the structure of a two zinc-finger DNA-binding domain from the protein Tramtrack complexed with DNA. The amino-terminal zinc-finger and its interaction with DNA illustrate several novel features. These include the use of a serine residue, which is semi-conserved and located outside the three key positions, to make a

base contact. Its role in base-recognition correlates with a large, local, protein-induced deformation of the DNA helix at a flexible A-T-A sequence and may give insight into previous mutagenesis experiments^{9,10}. It is apparent from this structure that zinc-finger/DNA recognition is more complex than was originally perceived.

Tramtrack (TTK) is a transcriptional regulator of the *Drosophila* development gene *fushi-tarazu*^{11,12}. The 66-residue DNA-binding domain (TTKDBD) contains two zinc-finger motifs with 10 residues (3 encoded by the vector) N-terminal to the conventional start of the first zinc-finger¹³. The crystal structure of the TTKDBD with an 18-base-pair (bp) oligonucleotide (Fig. 1a) was solved to 2.8 Å by isomorphous replacement (Table 1). The two zinc-finger motifs form independent DNA-binding modules (Fig. 1b), each positioned with the N terminus of an α -helix pointing into the major groove and making base-specific contacts. Most of the protein-DNA contacts are made to one strand of the binding site (Fig. 1a). The binding sites for the two fingers overlap. The alignment of the protein on its binding site and a number of the amino-acid to base contacts agree with our predictions based on footprinting analyses¹³.

Although the structure of finger 1 of TTK (and its interaction with DNA) shows several unanticipated features (see below), the structure of finger 2 is essentially the same as seen previously for other zinc-fingers^{4,14–16} (see legend to Fig. 1). Furthermore, finger 2 binds to DNA in a very similar manner to each of the three fingers of Zif268, with residues in equivalent positions making similar base and phosphate contacts (Fig. 3). Arg 152 and Asn 155 each make bidentate contacts to G8 and A7, respec-

TABLE 1 Statistics from the crystallographic analysis

Position of thymine substituted	Native	16	5	28
Resolution	2.8 Å	2.8 Å	3.2 Å	3.1 Å
Reflections total/unique	43,552/11,994	42,582/11,775	25,494/7,946	24,762/7,252
Multiplicity	3.6	3.6	3.2	3.4
Completeness	99.2%	99.3%	98.8%	82.4%
R_{merge}	5.6%	5.1%	8.7%	9.8%
	(2.87–2.8 Å 31.2%)	(2.87–2.8 Å 33.2%)	(3.37–3.2 Å 13.5%)	(3.26–3.1 Å 22.3%)
F/σ	41.9			
	(2.83–2.8 Å 9.65)			
Mean fractional isomorphous difference (3.3–3.2 Å)		6.4 (9.1)	11.4 (15.2)	13.6 (18.3)
MIR analysis 20–3.2 Å (3.57–3.24 Å):				
Phasing power	centric	0.5 (0.3)	0.8 (0.5)	0.5 (0.2)
	acentric	0.8 (0.6)	1.1 (0.8)	0.7 (0.5)
R_{Cullis}		0.77 (0.97)	0.67 (0.89)	0.83 (0.95)
Mean overall	centric	0.69 (0.50)		
figure of merit	acentric	0.46 (0.28)		

The protein and DNA used to obtain crystals are shown in Fig. 1a. The purification of the protein, the identification of an oligonucleotide sequence recognized specifically by the TTKDBD and the preparation of complex have been described previously¹³. For crystallization trials, a 1:1 complex was concentrated from ~15 μM to 0.3 mM. Crystals were grown in small tubes at 20 °C with a final complex concentration of between 0.075 and 0.15 mM in a buffer containing 20 mM MES pH 6, 5–20 mM NaCl and 1.75–3.5 mM spermine. Ordered native crystals could be grown from solution only once and further native and derivative crystals (using 5-bromouridine-substituted oligonucleotides) were grown by micro-seeding from one of the original crystals. Crystals grew to a maximum size of $200 \times 400 \times 600 \mu\text{m}$. They were stabilized by the addition of MPD (2-methyl-2,4-pentanediol) to 2% before mounting for data collection, and diffract to 2.8 Å at room temperature. The crystals are in the space group $P2_12_12_1$ with cell dimensions of $60.7 \times 64.6 \times 117.6 \text{ Å}$. The asymmetric unit contains two molecules of complex (two protein molecules and two DNA duplexes). Native and derivative data were collected using MAR (Hamburg, Germany) image plate scanners, both in our laboratory and at the Daresbury synchrotron source (station 9.5). These data were processed using IPMOSFLM²¹, merged using the CCP4 suite of programs²² and phases were calculated using MLPHARE²³. A solvent-flattened²⁴ MIR electron density map at 3.2 Å showed clear density for the DNA and 90% of the protein. A starting model of the complex based on the Zif268 structure and DNase I footprinting data for Tramtrack¹³ made map fitting straightforward. This was done using the graphics program O²⁵. The DNA was globally positioned in the map based on the coordinates of the bromine atoms in the derivatives and then individual bases were repositioned to give a better fit. The protein was positioned globally in a similar way by locating the high-intensity density for the zinc atoms and aligning the secondary structure in the model with the skeletonized density produced by the program BONES²⁶ implemented in O²⁵. The DNA molecules are packed end to end in the crystals by forming base triples: the terminal overhanging pyrimidine in one duplex makes a Hoogsteen interaction with the purine of the terminal base pair of the next duplex. The model was initially refined using X-PLOR²⁷. In early rounds the DNA was strongly constrained to maintain base planarity and pairing using the parameter file parallhdg.dna, but these constraints were relaxed towards the end of the refinement using the parameter file param11x.dna (ref. 27). Multiple rounds of rebuilding and refinement in X-PLOR were followed by refinement using TNT²⁸. Certain residues or side-chains were found to be poorly ordered in one or other or both of the complexes and have been omitted from the final model. These include: Met 101, Glu 102, Arg 114, Ser 117, Asn 137, Lys 165 and Ile 166. The current model contains 63 water molecules and has a R factor of 19.8% with r.m.s. deviation from ideal geometry of 0.015 Å in bond lengths and 2.14° in bond angles. The r.m.s. deviation between the B values of covalently bonded atomic pairs is 2.75 Å².

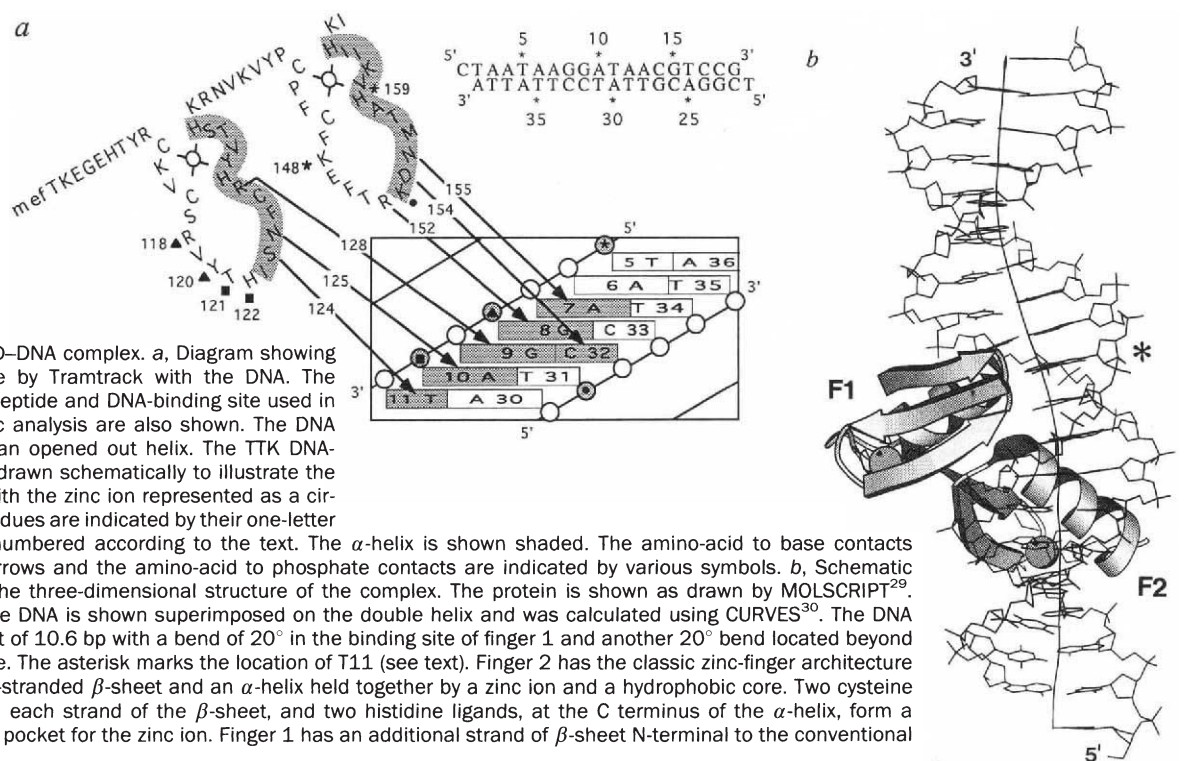


FIG. 1 The TTKDBD–DNA complex. *a*, Diagram showing the contacts made by Tramtrack with the DNA. The sequences of the peptide and DNA-binding site used in the crystallographic analysis are also shown. The DNA is represented as an opened out helix. The TTK DNA-binding domain is drawn schematically to illustrate the zinc-finger motif, with the zinc ion represented as a circle. Amino-acid residues are indicated by their one-letter abbreviation and numbered according to the text. The α -helix is shown shaded. The amino-acid to base contacts are indicated by arrows and the amino-acid to phosphate contacts are indicated by various symbols. *b*, Schematic representation of the three-dimensional structure of the complex. The protein is shown as drawn by MOLSCRIPT²⁹. The helix axis of the DNA is shown superimposed on the double helix and was calculated using CURVES³⁰. The DNA has a helical repeat of 10.6 bp with a bend of 20° in the binding site of finger 1 and another 20° bend located beyond the TTK binding site. The asterisk marks the location of T11 (see text). Finger 2 has the classic zinc-finger architecture consisting of a two-stranded β -sheet and an α -helix held together by a zinc ion and a hydrophobic core. Two cysteine ligands, located on each strand of the β -sheet, and two histidine ligands, at the C terminus of the α -helix, form a tetrahedral binding pocket for the zinc ion. Finger 1 has an additional strand of β -sheet N-terminal to the conventional zinc-finger.

tively. Asp 154 accepts two buttressing hydrogen bonds from Arg 152 and makes a good hydrogen bond to C32 on the opposite strand of the DNA. Such a contact was seen for finger 2 of Zif268, but had an unfavourable geometry. His 159, one of the zinc ligands, and Lys 148, from the second β -strand, both contact the phosphate 5P. Lys 153 in the α -helix contacts a phosphate (31P) across the major groove (Fig. 1a).

The residues linking the two finger domains have a high average temperature factor, indicating that this region is rather flexible or disordered. Hence the orientation of each finger domain is determined by the protein-DNA contacts and is unlikely to be influenced by the structure of the linker.

The structure of finger 1 shows that the residues N-terminal to the conventional finger motif fold to form a third strand to the β -sheet (Fig. 2a), in a similar way to the first zinc-finger of the protein SW15 (ref. 17). The additional β -strand, which is required for binding to DNA, is not seen to interact directly with DNA and is thus probably required for the structural stability of finger 1.

Several features of the interaction of finger 1 with DNA were unexpected. First, the zinc-histidine-phosphate 'charge relay' contact, observed in all three fingers of Zif268 and finger 2 of TTK, is substituted in finger 1 by a tyrosine-phosphate contact (Tyr 120 to 8P) (Fig. 2b). This variation is particularly interesting because the metal-binding histidine is invariant and had been assumed to serve as a phosphate anchor in all zinc-finger domains. The β -strand N-terminal to the α -helix makes many more contacts to the phosphate backbone than had been seen before (Arg 118 to 8P, Tyr 120 to 8P, Thr 121 to 10P and His 122 to 10P) and this is probably a consequence of a bend in the DNA (see below).

Second, the DNA in the crystal is deformed with a bend of 20° towards the protein in the binding site of finger 1 (Fig. 1b). This bend occurs at the readily deformable sequence A-T-A (ref. 18) (A10-T11-A12). The A-T and T-A steps have twists of 24° and 40° , respectively. Although alternate low and high twist are characteristic of this sequence¹⁹, the twist of the A-T step is particularly low and results in the bases of T11 and A10 stacking

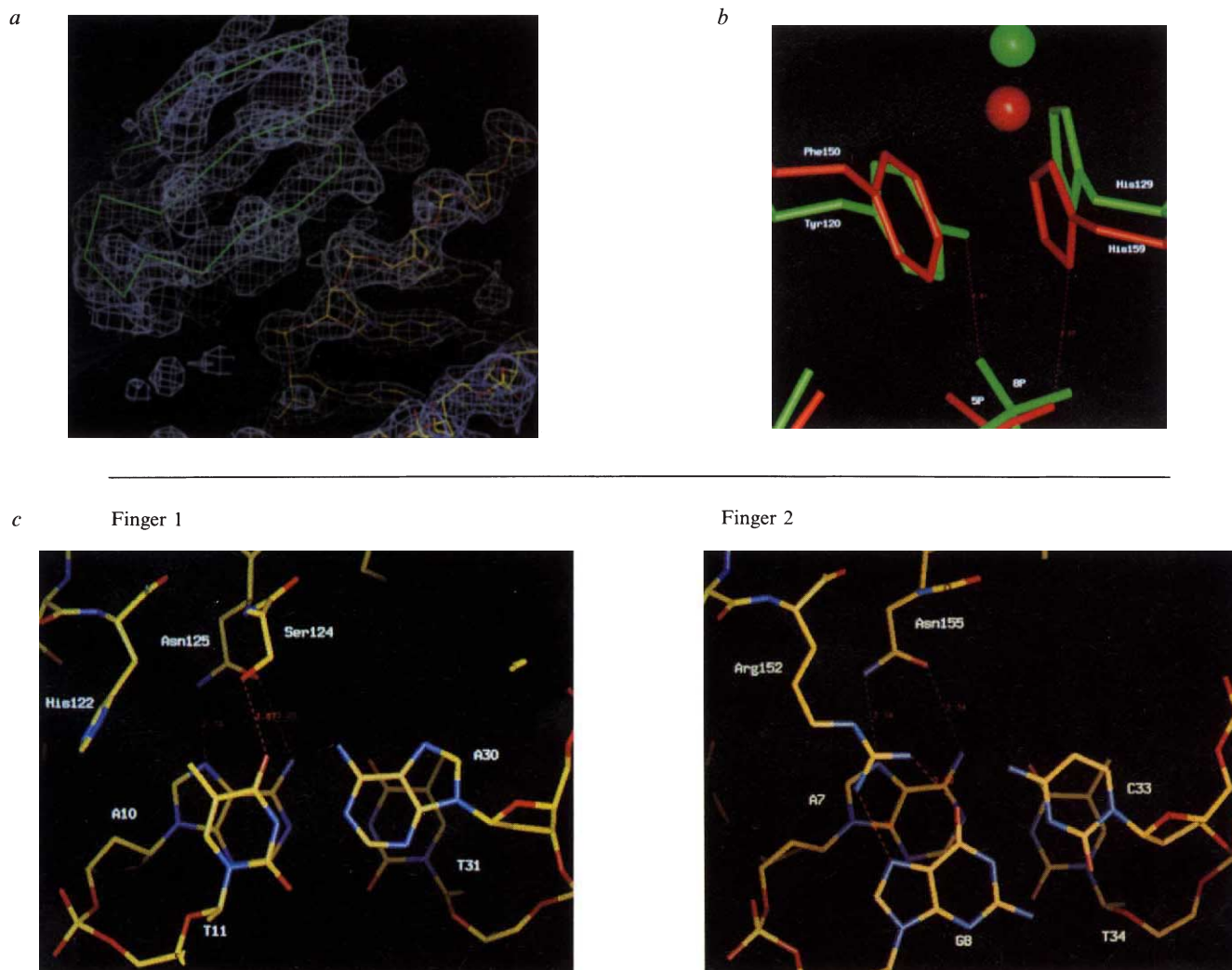


FIG. 2 Finger 1 has several features that are different from finger 2. a, The three β -strands of finger 1 shown in the $2F_o - F_c$ electron density map contoured at 0.3 electrons $\text{\AA}^{-3}$. This view also shows the phosphate backbones of the DNA and some of the base pairs in the minor groove. b, The histidine-phosphate contact observed in finger 2 (red), and the 3 fingers of Zif268, is substituted by a tyrosine-phosphate contact in

finger 1 (green). c, The distortion of the DNA helix (shown in Fig. 1) results in a displacement of T11 in the binding site of finger 1 towards the protein so that the base is within hydrogen-bonding distance of Ser 124. The amino-acid to base contacts for corresponding positions in the binding sites of fingers 1 and 2 are shown.

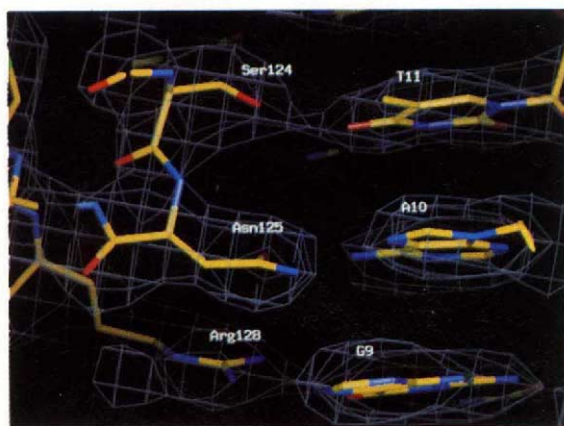
directly over each other. Consequently, T11 is displaced towards the protein by 2.5 Å compared with the equivalent base in the binding site for finger 2 (Fig. 2c). This permits the short side-chain of Ser 124 to interact with the O4 of T11. The distortion at base pair T11-A30 appears to be due to protein binding because in solution the N3 of A30 is reactive to dimethyl sulphate, but only when bound to protein¹³. This deformation illustrates that DNA need not be a passive partner in the recognition process¹⁸.

Finally, the pattern of amino acid/base contacts observed for finger 1 of TTK differs from that seen for finger 2 and the three fingers of Zif268 (Fig. 3c). A striking feature of the Zif268 structure was that it implied a general mechanism for zinc-finger/DNA recognition. In each finger, amino acids at three key positions (termed -1, 3 and 6 in Fig. 3c) have the potential to make base contacts to one strand of the DNA. The residues at positions -1 and either 3 or 6 make contacts to bases at

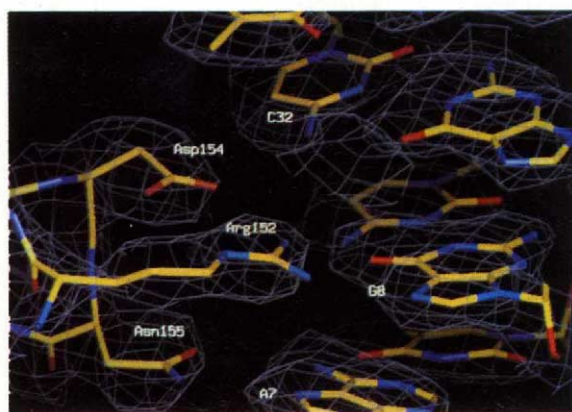
positions A and either B or C within a base triplet, respectively. This general mechanism was supported by a statistical analysis of a large number of zinc-finger sequences². Consequently, residues at these three key positions have been the focus of mutagenesis studies aimed at deriving a set of general rules for zinc-finger/DNA recognition⁶⁻⁸.

Finger 1 of TTK, unlike finger 2, does not conform to these simple rules (Fig. 3c). The histidine at position -1 is used to make a phosphate, rather than a base contact. Residues at key positions 3 and 6 make base contacts: Asn 125 to A10 (position 3 to B) and Arg 128 to G9 (position 6 to C) (Fig. 3a). Ser 124, in an additional position, 2, makes a third base contact to T11 at position A. Notably, serine is conserved at this position in over 50% of zinc-finger sequences². In several of the TTK binding sites the thymine T11 is substituted by a cytosine^{11,12,20}.

a



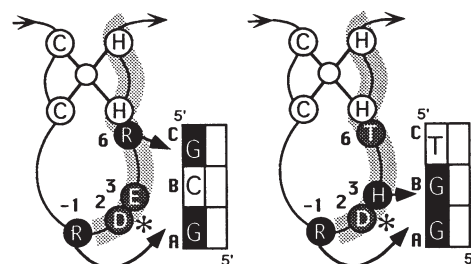
b



c

Zif268 F1 and F3

Zif268 F2



TTK F1

TTK F2

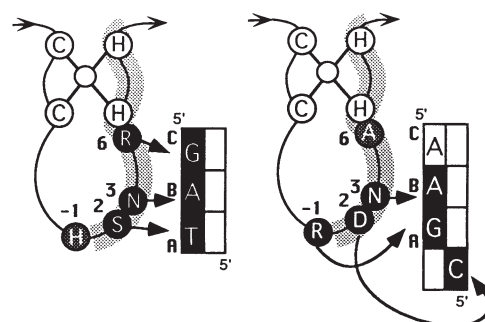


FIG. 3 The base-specific contacts made by the two zinc-fingers of Tram-track. **a**, Base contacts of finger 1 shown in the $2F_o - F_c$ electron density map contoured at 0.3 electrons Å⁻³. From top to bottom Ser 124 contacts T11, Asp 125 contacts A10 and Arg 128 contacts G9. **b**, Base contacts of finger 2 shown in the $2F_o - F_c$ electron density map contoured at 0.3 electrons Å⁻³. From top to bottom Asp 154 buttresses Arg 152 and contacts C32, Arg 152 contacts G8 and Asn 155 contacts A7. All the base contacts for finger 1 and finger 2 are made to one strand of the DNA with the exception of the contact to C32. **c**, Patterns of contacts in the three fingers of Zif268 (F1, F2 and F3) and the two fingers of TTK (F1 and F2) to their respective base triplets. The amino-

acid to base contacts are shown as white letters on a black background with arrows joining them. The four positions in the protein used for making base contacts are labelled -1, 2, 3 and 6 with the positions not used shown on a dark grey background. The positions are numbered according to the start of the α -helix. The three positions of the base triplet are labelled A, B and C. The location of the α -helix is indicated by the light grey shaded area. The asterisks mark the aspartic acids which make contacts to a base on the other strand of the DNA for the Zif268 fingers but were considered unimportant as described in the text.

Presumably a rotation about the C β -O γ bond would allow the serine to accept a hydrogen bond from N4 of the cytosine with no overall change in the geometry of the interaction. Substitution of the thymine by a cytosine, but not by a purine, should permit the DNA deformation described above. Hence it may emerge that fingers with a serine at position 2 have the potential to recognize a pyrimidine in position A, but this would probably require a deformation of the DNA.

It has been proposed from mutagenesis data that a glutamine in position -1 may recognize a thymine in position A (refs 9, 10). But to obtain such recognition, it was necessary to mutate additional residues: position 2 to a serine and position 3 to an aspartic acid^{9,10}. In the light of the TTK structure, we suggest that an alternative interpretation is possible: the serine, and not the glutamine, is contacting the thymine, and the glutamine would then be in a suitable position to make a phosphate contact. This alternative interpretation is supported by Nardelli *et al.* who had difficulty modelling the proposed glutamine to thymine contact because they found that only a long side-chain at position -1, such as that of arginine, could reach the DNA bases⁸.

In conclusion, the TTK structure reveals an extension to the pattern of direct protein-DNA contacts derived from the Zif268 structure. It remains to be seen whether further structures of zinc-finger/DNA complexes will reveal additional rules for zinc-finger DNA recognition. Establishing such rules is important if we are to design successfully zinc-finger proteins to recognize predetermined DNA sequences.

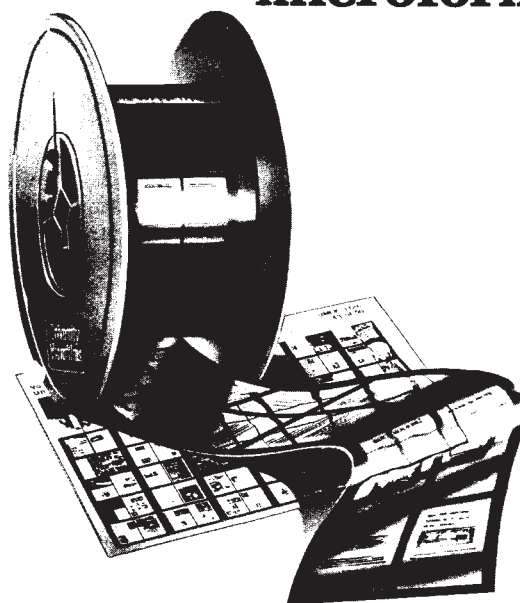
Finally, a recent structure determination³¹ supports our conclusions that not all zinc-fingers exhibit the same pattern of protein-DNA contacts seen for Zif268. The differences in this structure are greater than those between TTK and Zif268, but because three of the five fingers in this structure make virtually no contacts to DNA, it may be that this protein is atypical. $\square$

Received 6 August; accepted 15 October 1993.

1. Miller, J., McLachlan, A. D. & Klug, A. *EMBO J.* **4**, 1609-1614 (1985).
2. Jacobs, G. H. *EMBO J.* **11**, 4507-4517 (1992).
3. Hoovers, J. M. N. *et al.* *Genomics* **12**, 254-263 (1992).
4. Pavletich, N. P. & Pabo, C. O. *Science* **252**, 809-817 (1991).
5. Berg, J. M. *Curr. Opin. struct. Biol.* **3**, 11-16 (1993).
6. Desjarlais, J. R. & Berg, J. M. *Proc. natn. Acad. Sci. U.S.A.* **89**, 7345-7349 (1992).
7. Desjarlais, J. R. & Berg, J. M. *Proc. natn. Acad. Sci. U.S.A.* **90**, 2256-2260 (1993).
8. Nardelli, J., Gibson, T. J. & Chamay, P. *Nucleic Acids Res.* **20**, 4137-4144 (1992).
9. Desjarlais, J. R. & Berg, J. M. *Proteins* **12**, 101-104 (1992).
10. Desjarlais, J. R. & Berg, J. M. *Proteins* **13**, 282 (1992).
11. Harrison, S. D. & Travers, A. A. *EMBO J.* **9**, 207-216 (1990).
12. Brown, J. L., Sonoda, S., Ueda, H., Scott, M. P. & Wu, C. *EMBO J.* **10**, 665-674 (1991).
13. Fairall, L., Harrison, S. D., Travers, A. A. & Rhodes, D. J. *molec. Biol.* **226**, 349-366 (1992).
14. Lee, M. S., Gippert, G. P., Soman, K. V., Case, D. A. & Wright, P. E. *Science* **245**, 635-637 (1989).
15. Kleit, R. E., Herriott, J. R. & Horvath, S. J. *Proteins* **7**, 215-226 (1990).
16. Omichinski, J. G., Clore, G. M., Appella, E., Sakaguchi, K. & Gronenborn, A. M. *Biochemistry* **29**, 9324-9334 (1990).
17. Neuhaus, D., Nakaseko, Y., Schwabe, J. W. R. & Klug, A. *J. molec. Biol.* **228**, 637-651 (1992).
18. Travers, A. A. & Klug, A. in *DNA Topology and its Biological Effects* (eds Cozzarelli, N. R. & Wang, J. C.) 57-106 (Cold Spring Harbor Laboratory Press, New York, 1990).
19. Yoon, C., Prive, G. G., Goodsell, D. S. & Dickerson, R. E. *Proc. natn. Acad. Sci. U.S.A.* **85**, 6332-6336 (1988).
20. Read, D. & Manley, J. L. *EMBO J.* **11**, 1035-1044 (1992).
21. Leslie, A. G. W. In *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* Number 26 (Daresbury Laboratory, Warrington, UK, 1992).
22. CCP4 The SERC (UK) Collaborative Computing Project No. 4, a suite of Programs for Protein Crystallography (Distributed from the Daresbury Laboratory, Warrington, UK, 1979).
23. Otwinowski, Z. in *Isomorphous Replacement and Anomalous Scattering* 80-86 (Daresbury Laboratory, Warrington, UK, 1991).
24. Leslie, A. G. W. *Acta crystallogr.* **A43**, 134-136 (1987).
25. Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, N. *Acta crystallogr.* **A47**, 110-119 (1991).
26. Greer, J. *Meth. Enzym.* **115**, 206-224 (1985).
27. Brünger, A. T. *X-PLOR v. 3.0 manual* (Yale Univ., New Haven, 1992).
28. Tronrud, D. E., Ten Eyck, L. F. & Matthews, B. W. *Acta crystallogr.* **A43**, 489-501 (1987).
29. Kraulis, P. J. *appl. Crystallogr.* **24**, 946-950 (1991).
30. Lavery, R. & Sklenar, H. *J. biomolec. struct. Dynam.* **6**, 655-667 (1989).
31. Pavletich, N. P. & Pabo, C. O. *Science* **261**, 1701-1707 (1993).

ACKNOWLEDGEMENTS. We thank P. Evans and A. Leslie for their advice on many aspects of data collection and processing, T. Horsnell and K. Henrick for their help in compiling TNT, T. Smith and J. Fogg for oligonucleotide synthesis and A. Klug, A. Travers, Y. Choo, J. Li and W. Scott for comments on this manuscript. This work was in part supported by the Human Frontier Science Program.

nature is available in microform.



University Microfilms

International reproduces this publication in microform: microfiche and 16mm or 35mm film. For information about this publication or any of the more than 13,000 titles we offer, complete and mail the coupon to: University Microfilms International, 300 N. Zeeb Road, Ann Arbor, MI 48106. Call us toll-free for an immediate response: 800-521-3044. Or call collect in Michigan, Alaska and Hawaii: 313-761-4700.

☐ Please send information about these titles:

Name _____

Company/Institution _____

Address _____

City _____

State _____ Zip _____

Phone () _____

University
Microfilms
International

Familial startle disease is among the oddest of disorders. The demonstration of a mutation in the inhibitory glycine receptor points towards a cause of this and similar perturbations.

WHEN startled, most people jump, a reflex which is presumably a primitive defence mechanism. It is, perhaps, surprising then that people in whom this reaction is exaggerated (as in patients with startle disease, otherwise known as hyperexplexia) are probably more at risk than those lacking the reflex. In this month's *Nature Genetics*, John Wasmuth and colleagues¹ report that in some hyperexplexia patients a single amino acid is altered in the inhibitory glycine receptor. The many scattered observations surrounding this and similar conditions make more sense in the light of this discovery.

Hereditary hyperexplexia (also called Kok disease) is present from birth, affected infants displaying severe muscle rigidity and an exaggerated startle reflex. The sustained muscle rigidity (hypertonia) gradually subsides over the first year of life, although sufferers often die during this time — hypertonia of the respiratory muscles may mean they stop breathing. Throughout life, the exaggerated startle reflex can be brought on by an unexpected noise or tactile stimulus. The reflex may be accompanied by the severe rigidity which in some cases can cause the patient to become as stiff as a board.

Hyperexplexia is not unique in this respect. Several heritable disorders manifest reflex responses of such seriousness that patients can cause harm to themselves and onlookers. Notable among them is a condition known as Jumping or, more completely, the Jumping Frenchmen of Maine. The condition, first observed in French-Canadian lumberjacks from the Moosehead Lake region of Maine in northern America, has a dramatic phenotype. An exaggerated startle reflex is seen but this is often followed by reflex speech or behaviour. In an address to the American Neurological Association² in 1878, George Beard explained these manifestations as follows: "If [an affected individual] was abruptly asked to strike another, he would do so without hesitation, even if it was his mother and he had an axe in his hand". Other reflex reactions include echolalia (repetition of speech), echopraxia (copying a movement or action) and, reminiscent of a characteristic of Tourette's syndrome, coprolalia (speaking obscenities).

Descriptions of Jumping Men are unknown outside the northeastern United States and Canada. But it is difficult to believe that the existence of so-called

Goosey individuals, mainly in southern parts of the United States, can be coincidence. In 1980 Hardison³ described numerous cases of people who had an exaggerated reflex reaction when startled (in particular, when goosed). They demonstrated a marked jump, echolalia, occasional coprolalia and reflex carrying-out-of-commands. Jumping and Goosey may therefore be allelic disorders (or indeed a varying phenotype within the same disorder).

This idea is pure conjecture. But it is lent anecdotal support by an episode, in the second half of the eighteenth century, in which many French Canadians were expelled from the Maine area because they refused to swear allegiance to the British Crown. They fled to the southern states where many of them settled, and where Goosey seems to be most prevalent.

That behavioural disorders such as these can be a consequence of genetic mutation is intriguing. Moreover, Wasmuth and colleagues' report comes as something of a breath of fresh air. In each of six (soon to be seven) recent cases in which the genetic defect associated with a neurological condition has been characterized, that defect has been expansion of an unstable triplet repeat sequence⁴. With a single amino-acid substitution in a subunit of the glycine receptor, hyperexplexia is clearly an exception; and the new work has other unusual characteristics.

Hyperexplexia patients can be successfully treated with clonazepam, a benzodiazepine known to enhance γ -aminobutyric acid (GABA) neurotransmission. The sequence and structure of the GABA receptor is very similar to those of the glycine and glutamate receptors. Using a standard linkage approach, Wasmuth and co-workers had previously uncovered linkage of hyperexplexia to the long arm of chromosome 5 and found the area to be uncommonly rich with good candidate genes, including those coding for two subunits of the GABA receptor (GABRA1 and GABRA2), a glutamate receptor (GLUR1) and a subunit of the glycine receptor (GLRA1). Spoilt for choice, the authors used radiation hybrid analysis to further localize these genes and were able to show that GABRA1 and GABRA2 fell outside the hyperexplexia candidate region whereas GLRA1 was within it. They also found that the glycine receptor is antagonized by strychnine

which, when administered to mice, causes hypertonia and an exaggerated startle reflex. So a good candidate gene became an excellent one and they started screening GLRA1 in affected patients and their families. Four of the seven families studied showed evidence for mutations, all of which involved the substitution of the same amino acid in the mature protein.

In an accompanying News and Views article⁵, Floeter and Hallett point out that the commonly held view within the neurological community, that hyperexplexia patients show a hyperexcitability within the reticular formation of the brain stem, does not tie in with the new findings, because the glycine receptor is found only at relatively low levels in the reticular formation. But they go on to explain that although the acoustic startle response in hyperexplexic patients is essentially normal, with patients just being more sensitive to sound and demonstrating a greater response to it, they do respond abnormally to other stimuli. Taps on the nose and experimental electrical stimulation of the peripheral nerves also lead to a tremendous response. But such responses clearly involve the central and peripheral nervous systems, which implies that hyperexplexia may be a generalized neural hyperexcitability of normal pathways (rather than being restricted to a certain area of the central nervous system), and that the startle response may simply be the most prominent manifestation of this generalized hyperexcitability.

This idea fits in much better with the findings of Wasmuth and colleagues. Glycine is known to inhibit signals of the nervous system by binding to chloride ion-channel receptor complexes, a process which takes place in at least two populations of glycinergic inhibitory interneurons connected to motor neurons of the central nervous system. Hence a defect in the glycine receptor would result in a removal of an inhibitory loop, and could therefore explain some of the characteristics of hyperexplexia and possibly other inherited startle disorders.

Adrian J. Ivinson

Adrian J. Ivinson is Assistant Editor of Nature Genetics.

1. Shiang, R. *et al. Nature Genet.* **5**, 351–358 (1993).
2. Beard, G.M. *J. nerv. ment. Dis.* **5**, 526 (1878).
3. Hardison, E. H. *J. Am. med. Ass.* **244**, 70 (1980).
4. Davies, K. *Nature* **364**, 88 (1993).
5. Floeter, M. K. & Hallett, M. *Nature Genet.* **5**, 319–320 (1993).

Taking a calculated opportunity

Douglas McCormick

The natural familiarity of graduate science students with computers and software opens up new opportunities in areas both related to science and well away from it, in the financial world.

In a world of laboratory information management systems and *in electro* experimentation, bit-flipping has become a fact of life that can be decried but not denied: database management, data analysis; e-mail, on-line journals, and 'collaboratories'; dynamic models of star formation, neural nets, fuzzy logic, artificial intelligence, and genetic-algorithm models of everything from economics to evolution; databanks of gene sequences and protein structures; molecular models and computer-aided regulatory submissions.

Saying that there are job opportunities for scientists with computer skills is like saying there are job opportunities for scientists who can read and write. Certainly, programming, systems, or communications knowledge gives a scientist, especially a junior researcher, wider options — both within a scientific discipline and, should need arise, beyond it.

According to the US Bureau of Labor Statistics (BLS)¹, the country employs some 144,000 chemists and chemistry professors, 127,000 life-scientists, 58,000 geologists and earth scientists, and 38,000 physicists. (European figures were not available, but are probably comparable.) Some 2.2 million Americans program computers for a living — electrical and electronics engineers (of whom about one in four works on computers), systems analysts, operations analysts, computer scientists, programmers, and operators. The commercial software business alone employs 330,000 people to program, produce, and market the programs that millions of others use daily in their jobs or at home. Six hundred thousand personal computers were sold in the USA last year alone.

The computing sector is both booming and contracting. Overall computing employment dropped 3.9 per cent from 1991 to the end of 1992.

Software-industry employment decreased by 3 per cent — thanks to massive layoffs at old-line computer and software companies and brutal price wars among publishers of personal computer software. Yet worldwide software sales increased by more than 10 per cent in 1992, making a \$52 billion global market. And new software companies raised about \$2 billion in expansion capital in 1992 — about as much as all the other electronic industry segments combined.

These trends are bound to accelerate. A report from Ernst & Young², the accounting and consulting firm, rather neatly summarizes the looming changes in the software business — and thus describes the pending working environment of those who use computers. These imminent critical shifts are, using the accountants' jargon:

- From technology push to customer pull.
- From vertical integration to virtual enterprise.
- From domestic focus to global outlook.
- From hardware-driven to software-driven.
- From proprietary systems to industry standards, integration, and open systems.
- From centralized mainframe processing to decentralized, networked PC processing.

Biology is destiny

Except for the biophysicists, who have long depended on computers for the

complex energy minimizations required for molecular modelling, biologists have seemingly been averse to using computers for anything but record keeping. While every second scientific exhibition stand shimmers in the glow of a new computer terminal, conference sessions on computer tools in biology are customarily empty. That, think some researchers and educators, is a situation that is due for a change.

What it boils down to, says Leroy E. Hood, chairman of the Department of Molecular Biotechnology at the University of Washington, is deciding what students should be taught, which leads quickly to the question of what computers will do for biology.

Hood's new department is built on the conviction that interdisciplinary education is the way to meet the complex challenges of scientific research. The group envisages graduate education that cross-fertilizes molecular biology with chemistry (bio- and otherwise), applied mathematics, applied physics, bioengineering, electrical engineering, and, not least, computer science.

Just as graduate students have traditionally been required to speak and write a foreign language, they should also know a programming language. It does not matter which language precisely (C, C++, and Pascal are common choices, though other tools might be acceptable). What is important is that the student master the fundamental conceptual tools.

At the same time, says Hood, every student should be able to understand and use competently the five or six algorithms most common in molecular biology research: multi-sequence DNA and peptide

TABLE 1 US employment in thousands, December 1992

	Total	Agriculture	Mining	Construction	Manufacturing	(Non-durable)	Transportation, media	Wholesale trade	Retail trade	Finance, insurance, real estate	Total miscellaneous services	Total professional services	(Hospital/s)	(Other health services)	(Educational)	(Social services)	(Other professional services)	Public administration
Electrical and electronic engineers	527	0	1	19	253	14	103	9	2	2	105	69	0	0	8	0	61	33
Computer systems analysts and Scientists	693	0	5	3	153	31	49	16	17	61	321	84	9	3	28	2	42	68
Operations and systems researchers	192	0	3	2	46	12	24	5	5	29	40	29	4	2	5	1	17	37
Computer science teachers	15	0	0	0	0	0	0	0	0	0	15	15	0	0	15	0	0	0
Computer programmers	549	1	8	4	128	30	48	17	25	68	209	67	9	5	26	1	25	43
Computer operators	660	2	4	8	142	64	50	39	46	106	202	120	25	15	33	7	40	60
Computing total	2636	3	21	36	722	151	274	86	95	266	892	384	47	25	115	11	185	241

Source: US Bureau of Labor Statistics.

TABLE 2 Average salary by job function and industry

	% Of Respondents	LAN manager/ administrator	Systems & programming manager	Senior programmer/ analyst	Programmer	Database manager/ administrator	Database analyst
Average salary		38,647	55,113	42,659	30,491	48,050	41,747
Average annual bonus		1,895	4,167	1,769	1,674	2,691	1,532
Total compensation		40,542	59,280	44,455	32,165	50,741	43,279
Banking	4.3%	36,357	56,642	45,458	31,083	44,500	39,000
Insurance	5.5%	40,781	58,718	45,360	32,333	51,656	43,650
Real estate	1.4%	38,500	66,500	38,000	34,000	#N/A	#N/A
Securities	1.8%	35,750	61,000	42,200	30,400	44,000	#N/A
Computer-related	1.4%	41,166	62,666	40,687	25,761	31,713	32,612
Manufacturing	28.0%	41,306	53,896	43,800	30,698	61,785	41,345
Media	2.1%	41,200	68,000	48,888	35,000	43,214	32,508
Retail/wholesale	14.0%	36,500	53,309	41,238	29,133	40,158	52,042
Transportation	2.3%	40,800	53,142	43,529	32,000	43,500	37,500
Utilities	2.9%	37,488	59,166	46,916	26,153	50,285	40,093
Business services	2.6%	34,250	54,156	40,679	26,102	42,191	47,625
Education	6.9%	32,961	43,900	35,826	28,107	43,437	36,166
Government	11.9%	40,240	54,437	42,550	32,050	48,155	46,815
Healthcare	6.7%	34,500	58,123	38,951	29,857	42,791	34,600
Non-Profit	1.6%	33,500	47,600	41,333	33,666	38,000	37,000

Source: *Computerworld*

alignment, genealogical trees, similarity searches, genetic-fingerprint analysis, gene mapping, and the like.

Databases present a great challenge to life scientists, in Hood's estimation. They should be easy to search, easy to link together, and clearly annotated in a way that illuminates the biological significance of the information. In these tasks, Hood says, we are clearly failing. His own group is working on a method of linking sequence and biology, annotating the sequences of the T-cell receptor β locus as a model. ("It's so complicated; if we can manage that, we can manage anything").

At present, conventional relational database systems only allow meaningful searches and associations in categories for which the database designer and programmer have made provision. Current databases are thus unresponsive to expanding knowledge and new schemes of categorization in evolving research disciplines.

This issue of flexibility leads naturally to issues of database transparency. Ideally, researchers connected to the data highway should be able to sift through any or all of the tens of thousands of balkanized scientific databases, using a unified interface and query language to ferret out fresh connections, free of today's towering technical obstacles. Thus, at the very least, the researcher should be able to link the one-dimensional mainstay of modern molecular biology, the DNA sequence library, with the three-dimensional protein structure databases...and work towards the four-dimensional, time-based structures that are now the domain of the developmental biologist, the complex systems theorist, and the neural-net experimenter.

Enterprise computing

Andy Lawton's career is the very model of interdisciplinary cross-fertilization, distributed processing, open systems, and global outlook.

Lawton began as a statistical geophysicist specialising in nonlinear optimizations. The statistical background led to consulting for research physicians. That, in turn, led to his present position, head of Boehringer Ingelheim's office of clinical data management, a small group based in Bracknell, UK, that coordinates data consolidation and analysis for more than 100 clinical trials programmes in the United States, Canada, France, Italy, Germany, The Netherlands, and elsewhere. This 'slide' across many disciplines is a boon, rather than a burden, Lawton says. Otherwise, "one gets too narrow and doesn't realize that the problems one faces have already been solved elsewhere."

The hallmarks of this new-generation approach, 'enterprise computing' in bit-speak, are openness, decentralization, and standardization. The Boehringer system started as an expedient to solve problems and save time. It could have produced the same sort of system favoured by many a big pharmaceutical company: a batch-fed mainframe tended by the management information systems department, dedicated to running payroll and accounting programs by day, and digesting clinical trials data at night.

Instead, Boehringer developed tools to link together people in the field in a standardized, company-wide, world-wide system that integrates the many aspects of clinical management into a single, comprehensible user interface. Such a system, Lawton says, more than meets the corporate goals of problem solving and efficiency.

Practical tests show that distributed desktop computing produces results about six times faster than a batch/mainframe system. The advantages go further, though: a well thought out enterprise system promotes communication within the company and can very well evolve from a cost centre into a profit centre, as the easy-to-use tools are made available to outside organizations.

Modern programming tools offer other advantages as well, Lawton notes. He likes OS/2 because it is a true multitasking operating system (and is robustly stable), allowing the heavy statistical number crunching to fade into the background, leaving the computer still able to function as a workstation for applications development, communications, or data entry.

Like other modern operating systems, OS/2 lends itself to object-oriented programming, the process of breaking large applications down into independent function libraries that the program taps during its run. Using object modules reduces development time. More to the point, it reduces regulatory delays, since applications developers reuse code already tested and validated by the national drug regulatory authorities.

Finally, says Lawton, graphic user interfaces (GUIs, a class that includes the Macintosh operating system, OS/2, Microsoft Windows, and X Windows) lend themselves to applications that follow the user's natural thought processes and work flow.

The key to quality, says Lawton, is to keep the team small and the accountability high. Boehringer's pioneering system was built by three people in conjunction with some 200 users, an example of what you can do if people know the area they are working in, Lawton says.

Lawton seeks the same sort of openness in co-workers that he favours in systems. Though Lawton does look for some programming or clinical trials experience in job applicants, "a nice attitude" can take precedence over an overstuffed curriculum vitae. Sometimes, indeed, new systems staff must forget what they may have learned in the past and start afresh on the problems. Too many people, in biology and computing, hide behind three-letter abbreviations.

Such open systems have certain incentives built in. Members of the Boehringer Ingelheim clinical trials staff plan, write, and support their applications. They also deal with users' problems and complaints, a system which leaves them with built-in motivation to make their programs bullet proof.

A good reward system is important. Starting salary is less important than continued raises to recompense good work. In the USA, on the other hand, salaries tend to start high, but then do not increase in line with improved performance. This is a shame, says Lawton, because it reduces the programmer's feeling of value. That is dangerous, because, "you can't afford to lose good people," and good computer people can work virtually anywhere, their skills as valuable to a City financial firm as to a pharmaceutical company.

Bailing out

Dick Lewis is an employment recruiter at Kaye-Tech, a New York City data-processing placement firm. His company follows the demand for programmers and analysts, and of late most of that demand has come not from scientific institutions but from Wall Street. A boom in activity in the financial markets, possibly brought on by falling interest rates, has produced a demand for quantitative thinkers and programmers that has lured a number of PhDs out of research and commercial research and development and into finance. Lewis, too, finds that scientists' computer skills translate easily to investment banking, portfolio management, and other financial services.

For their part, PhDs leave the academic world for higher salaries, a higher profile, and the vertiginous challenges of Wall Street.

Financial houses are particularly interested in researchers from disciplines rich in mathematics. The case of a Columbia assistant professor of physical sciences is illustrative. The erstwhile academic found a job at a company that supplies financial research for brokerage houses. While established brokerages still tend to prefer candidates with financial backgrounds, contract vendors are hungry for people who are able to think independently and adapt.

Comfort with mathematical reasoning



is critical. The trend is away from the sort of programmer who knows code inside-out but nothing about the business he or she is writing for. The emerging breed combines programming with operations research: increasingly, employers are looking for people who can think through complex interrelationships, reduce them to mathematical models, and express those models in usable code, Lewis says.

Lewis places about 300 candidates a year; 10 to 15% of those candidates come from scientific backgrounds and have no previous experience in finance. Lewis has been seeing increasing number of applications from programmer-analysts in big private laboratories like AT&T Bell Laboratories and the IBM Thomas J. Watson Laboratories, many also with significant research backgrounds.

C++ has emerged as the lingua franca of programming. It is (by derivation, anyway) Unix's native tongue, the darling of operating system developers generally, and the language of choice for writing applications to run under Windows or OS/2.

Wall Street is part of the trend. Financial employers are typically looking for people who can write C++ programs to run under Unix on Sun workstations, Lewis says. Other sought-after skills include Unix, VMS, MS-DOS, Windows, relational database management, and Standard Query Language facility.

Compensation and opportunity

The BLS compiles employment profiles of the working populace through a survey of 60,000 households, 110,000 to 120,000 individuals. The data for 1992 show some 2.6 million Americans employed in computing and electronic engineering (Table

1), a total 3.9 per cent lower than that for 1991. Those results tally with data compiled in an Institute of Electrical and Electronics Engineers (IEEE) study³ released last May, which found employment levels 3 per cent down on the previous year. A BLS spokesman noted, however, that employment levels in the USA have increased this year, especially in the early months, during which growth in service jobs more than compensated for continued erosion in manufacturing.

Interestingly, the BLS figures show total employment among life scientists and university biology teachers also down about 3 per cent for the period, though this loss is more than outweighed by gains in employment for biological technicians, medical scientists, and physicians.

The BLS reports that job growth in computing disciplines is largely limited to manufacturers of non-durable goods (a sector that includes the pharmaceutical industry, publishing, and chemicals) and professional services (hospitals, doctors' offices, research and development, engineering and the like).

In its annual salary survey, *Computerworld* magazine analysed 1,239 computer professionals' income by job title, industry, region, and company size⁴. Selected data (Table 2) show that local area network managers earn an average \$38,647 a year, senior programmer/analysts make \$42,659 a year, programmers earn \$30,491, database administrators make \$48,050, and database analysts make \$41,747.

Computer engineers comprised by far the largest primary area of technical competence among the 4,282 engineers who responded to the IEEE study. Twenty-three per cent of them worked on software (commonly on software alone, but sometimes in conjunction with hardware development). This group earned a median \$60,000 a year, ranging from \$38,000 in the bottom decile to \$103,000 in the top.

Douglas McCormick is a freelance writer based at 145 Blauvelt Street, Teaneck, New Jersey 07666, USA. He was until recently editor of Bio/Technology magazine and vice president/editorial director of Nature America. He was formerly computer science editor of Hayden Book Company.

1. *Industry and Occupation Tables*. Table 5. Employed Persons by Detailed Occupation and Major Industry. Bureau of Labor Statistics, US Department of Labor, Washington, DC.

2. Burrill G.S. & Almasy S.E., *Electronics 93: The New Global Reality*, Ernst & Young, San Francisco, 1993.

3. *IEEE US Membership Salary and Fringe Benefit Survey 1993*, (Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, 1993).

4. Wilde C. *Computerworld*, 6 September 1993, 91-100.

PLEASE follow these guidelines so that your manuscript may be handled expeditiously.

Nature is an international journal covering all the sciences. Contributors should therefore bear in mind those readers who work in other fields and those for whom English is a second language, and write clearly and simply, avoiding unnecessary technical terminology. Space in the journal is limited, making competition for publication severe. Brevity is highly valued. One printed page of *Nature*, without display items, contains about 1,300 words.

Manuscripts are selected for publication according to editorial assessment of their suitability and reports from independent referees. They can be sent to London or Washington and should be addressed to the Editor. Manuscripts may be dealt with in either office, depending on the subject matter, and will where necessary be sent between offices by overnight courier. High priority cannot be given to pre-submission enquiries; in urgent cases they can be made in the form of a one-page fax. All manuscripts are acknowledged on receipt but fewer than half are sent for review. Those that are not reviewed are returned as rapidly as possible so that they may be submitted elsewhere without delay. Contributors may suggest reviewers; limited requests for the exclusion of specific reviewers are usually heeded. Manuscripts are usually sent to two or three reviewers, chosen for their expertise rather than their geographical location. Manuscripts accepted for publication are processed from the London office.

Nature requests authors to deposit sequence and crystallographic data in the databases that exist for this purpose.

Once a manuscript is accepted for publication, contributors will receive proofs in about 4 weeks. *Nature's* staff will edit manuscripts with a view to brevity and clarity, so contributors should check proofs carefully. Manuscripts are generally published 2–3 weeks after receipt of corrected proofs. *Nature* does not exact page charges. Contributors receive a reprint order form with their proofs; reprint orders are processed after the manuscript is published and payment received.

Categories of paper

Review Articles survey recent developments in a field. Most are commissioned but suggestions are welcome in the form of a one-page synopsis addressed to the Reviews Coordinator. Length is negotiable in advance.

Progress articles review particularly topical developments for a nonspecialist readership. They do not exceed 4 pages in length. Suggestions may be made to the Reviews Coordinator in the form of a brief synopsis.

Articles are research reports whose conclusions are of general interest and which are sufficiently rounded to be a substantial advance in understanding. They should not have more than 3,000 words of text (not including figure legends) or more than six display items and should not occupy more than five pages of *Nature*.

Articles start with a heading of 50–80 words written to advertise their content in general terms, to which editors will pay particular attention. The heading does not usually contain numbers, abbreviations or measurements. The introduction to the study is contained in the first two or three paragraphs of the article, which also briefly summarize its results and implications. Articles have fewer than 50 references and may contain a few short subheadings.

Letters to Nature are short reports of outstanding novel findings whose implications are general and important enough to be of interest to those outside the field. Letters should have 1,000 or fewer words of text and four or fewer display items. The first paragraph describes, in not more than 150 words and without the use of abbreviations, the background, rationale and chief conclusions of the study for the particular benefit of non-specialist readers. Letters do not have subheadings and contain fewer than 30 references.

Commentary articles deal with issues in, or arising from, research that are also of interest to readers outside research.

News and Views articles inform nonspecialist readers about new scientific advances, sometimes in the form of a conference report. Most are commissioned but proposals can be made in advance to the News and Views Editor.

Scientific Correspondence is for discussion of topical scientific matters, including those published in *Nature*, and for miscellaneous contributions. Priority is given to letters of fewer than 500 words.

Preparation of manuscripts

All manuscripts should be typed, double-spaced, on one side of the paper only. An original and four copies are required, each accompanied by artwork. If photographs are included, five sets of originals are required; for line drawings, one set of originals and four good-quality photocopies are acceptable. Reference lists, figure legends and tables should all be on separate sheets, all of which should be double-spaced and numbered. Three copies of relevant manuscripts in press or submitted for publication elsewhere should be included with submitted manuscripts, clearly marked as such. Five copies of revised and resubmitted manuscripts, labelled with their manuscript numbers, are required, together with five copies of a letter detailing the changes made.

Titles are brief and simple. Active verbs, numerical values, abbreviations and punctuation are to be avoided. Titles should contain one or two key words for indexing purposes.

Artwork should be marked individually and clearly with the author's name and, when known, the manuscript number. Ideally, no figure should be larger than 28 by 22 cm. Figures with several parts are to be avoided and are permitted only if the parts are closely related, either experimentally or logically. Unlettered originals of photographs should be provided. Suggestions for cover illustrations, with captions and labelled with the manuscript number, are welcome. Original artwork is returned when a manuscript cannot be published.

Protein/nucleotide sequences should ideally be in the three-letter and not the single-letter code for amino acids. One column width of *Nature* can accommodate 20 amino acids or 60 base pairs.

Colour artwork. A charge of £500 per page is made as a contribution towards the cost of reproducing colour figures. Inability to pay these costs will not prevent publication of essential colour figures if the circumstances are explained. Proofs of colour artwork may be sent to contributors under separate cover from their galley proofs.

Figure legends should not exceed 300 words and ideally should be shorter. The figure is described first, then, briefly, the method. Reference to a method published elsewhere is preferable to a full description. Methods are not described in the text.

References are numbered sequentially as they appear in the text, followed by those in tables and finally by those in figure legends. Only papers published or in the press are numbered and included in the reference list. All other forms of reference should be cited in the text as a personal communication, manuscript submitted or in preparation. Text is not included in reference lists. References are abbreviated according to the *World List of Scientific Periodicals* (Butterworths, London, 1963–65). The first and last page numbers are included; reference to books includes publisher, place and date.

Abbreviations, symbols, units and Greek letters should be identified the first time they are used. Acronyms should be avoided whenever possible and, if used, defined. Footnotes are not used in the text.

Acknowledgements are brief and appear after the reference list; grant and contribution numbers are not allowed.

Supplementary information is material relevant to Articles or Letters which cannot, for lack of space, be published in full, but which is available from *Nature* on request.

Submission. Manuscripts can be sent to the Editor at 4 Little Essex Street, London WC2R 3LF, UK or at 1234 National Press Building, Washington, DC 20045, USA. A telephone and fax number should be included. Manuscripts or proofs sent by air courier to London should be declared as 'manuscripts' and 'value \$5' to prevent the imposition of import duty and value-added tax.